

# nature

## NEEDLE IN A HAYSTACK

Tracking down the fragments  
of asteroid 2006 TC<sub>3</sub>

**GENE PATENTS**

The jury is still out

**VOLCANIC PUMICES**

What they reveal

**AQUACULTURE**

Keep fish on the menu

# Embryonic education

Now that the US federal funding ban on human embryonic stem cells is lifted, scientists must engage the public's concerns about embryo research.

When US President Barack Obama lifted the funding ban for research on human embryonic stem cells earlier this month, he did not mention the Dickey–Wicker amendment — legislation that forbids the use of federal funds for research that destroys or creates embryos. It was a missed opportunity to begin a necessary conversation.

Because of this law, worthy projects will still be barred from federal funding despite Obama's action. Stanford University, for example, collaborates with fertility clinics that work with couples who have their embryos screened for genetic diseases. If pre-implantation embryos are identified with genetic mutations, the prospective parents can donate them to the research project, allowing stem-cell lines to be derived from them. This way, embryos unsuitable for implantation would become cell lines for studying diseases afflicting the donors' families. Although the programme has California state funding, this does not cover the costs of deriving lines from all the embryos donated, which in turn restricts the research.

In force since 1996, the Dickey–Wicker amendment badly needs updating to fit the current research reality, if not outright repeal. But because it affects fewer researchers than did the funding restrictions on stem-cell research, scientists who spent hours in public outreach trying to overturn the stem-cell ban may well want to return to their labs, leaving this lower-profile law's implications unquestioned.

Such attitudes are understandable, but wrong. Both the Dickey–Wicker amendment and the new guidelines on human embryonic stem-cell research being drawn up by the National Institutes of Health merit an intense national conversation. In particular, that dialogue should thoroughly explore attitudes towards studying different types of embryos — not just those left over from fertility procedures, but also those that might be specially created for research.

The United Kingdom set a good example. More than 25 years ago, the government began supporting a series of public dialogues about what sorts of embryo research would be deemed acceptable. This helped breed a trust and openness between the general public and

the scientific communities that has permitted the scope of allowable research to expand over time.

In the United States, scientists should likewise highlight the restrictions they have already imposed on themselves — especially their use of ethical oversight committees, which bring together scientists, members of the local community and ethical expertise to ensure that research falls within established guidelines and has scientific merit. Scientists should also describe how they balance the status of human embryos with the potential benefits of research. And they should listen carefully to non-scientists' objections, hopes and concerns.

A key requirement for productive dialogue is a common frame of reference. Here, the word 'embryo' is a stumbling block. This term refers to everything from a newly fertilized single-celled egg to millions of cells organized into eyelids, ears, genitals and limbs. Yet the latter form, which is present some eight weeks after fertilization, is not only ethically unacceptable for research but also far too old to yield embryonic stem cells.

Multiple sets of widely accepted guidelines from, for example, the US National Academies, the International Society for Stem Cell Research and Britain's Warnock Report agree that the first sign that cells for the future body are starting to specialize — the glimmer of a structure known as the primitive streak at about 14 days after an egg begins to divide — marks the end of when any laboratory research on human embryos should be considered. To discuss this responsibly, scientists should insist on precision, specifying an embryo's developmental state in terms of its age, for example, or the number of cells.

In the United Kingdom, a broad consensus on stem-cell research began with long, ongoing consultations with the public. That discussion — using the appropriate language — must now begin in earnest in the United States. ■

**"Scientists should describe how they balance the status of human embryos with the potential benefits of research."**

## A cut too far

UK researchers are rightly outraged at one funding council's decision to exclude certain applicants.

For around 250 British scientists, April's post will bring a particularly personal letter of rejection. The Engineering and Physical Sciences Research Council (EPSRC), a national research-funding body, has decided to stop serially unsuccessful applicants from submitting any more grant proposals for a year (see page 391).

No other funding body, whether in the United Kingdom, the rest

of Europe or the United States has attempted to formally exclude scientists in this way. The anger triggered by the move tells a cautionary tale: the policy is misguided, however urgently the EPSRC needs to relieve its overburdened peer-review system.

In truth, the research council is caught in a vicious circle. Its budget for grants shrank this year as government funding increases were swallowed up to pay for the full economic costs of research. Yet the applications keep flooding in. The result is that only one in every four or five applications is currently being funded; in some disciplines, such as chemistry, success rates have dropped to 15%. In the hope of getting at least some support, researchers feel compelled to submit more and more applications for small, short-term grants,



## RESEARCH HIGHLIGHTS

**Deeper still and deeper**

*Proc. R. Soc. B* doi:10.1098/rspb.2009.0102 (2009)

Darwin suggested that particularly long-tongued pollinators and long-tubed flowers are dancers in a sort of coevolutionary tango, each acting as selection pressure on the other. When faced with a flower that keeps its nectar at the end of a long tube, a nectar bat (*Anoura fistulata*, pictured) with a longer-than-average tongue has an advantage over its shorter-tongued conspecifics. Now, Nathan Muchhala and James Thomson of the University of Toronto, Canada, have shown that pressures run the other way as well.

The duo experimentally manipulated the length of the flower *Centropogon nigricans* and found that longer tubes delivered more pollen to bats in male flowers, and took more pollen from bats in female flowers. The exact mechanism, however, remains unknown.



N. MUCHHALA & J.D. THOMSON

**MARINE BIOLOGY****Snotty sampling**

*Mar. Mamm. Sci.* doi:10.1111/j.1748-7692.2008.00277.x (2009)

Water ejected from whale blow holes can be used for research into the animals' hormonal state, and possibly gender.

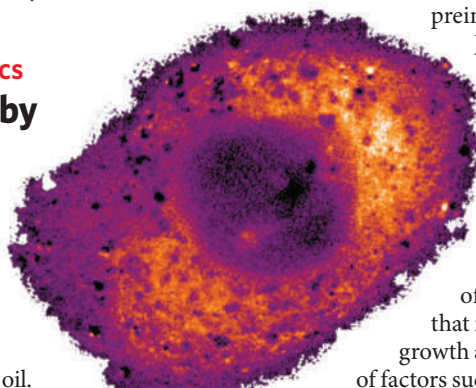
Researchers studying terrestrial mammals can measure hormone levels in faeces and urine; with aquatic mammals, such non-invasive monitoring is difficult. Using nylon stockings mounted on 13-metre carbon-fibre poles, Carolyn Hogg of the University of New South Wales in Sydney, Australia, and her colleagues sampled the exhalations of humpback and northern right whales they were following to see whether these might provide a source of hormonal data.

The team discovered progesterone and testosterone in some of the blows. Testosterone collected from single whales escorting female–calf pods supports theories that escorts are males waiting to breed. However, further testosterone/oestradiol tests are needed to definitively determine the whales' sex.

**MOLECULAR DIAGNOSTICS****Like a record, baby**

*Nature Chem.* doi:10.1038/nchem.120 (2009)

The fluids in the different compartments inside cells vary in viscosity, with those in some parts as thin as water, and those in others more glutinous than olive oil.



Marina Kuimova of Imperial College London and her colleagues have now watched as a dying cell's fluids thicken. Their tool, and weapon, was a molecular rotor that gives off a fluorescent signal that varies according to its rotation rate and can be triggered with light to release reactive oxygen species, which harm the cell.

When the lights went on and the harmful oxygen species were released, the spin rates of the molecules slowed. The team thinks that the cell's internal fluids became drastically more viscous (shown as orange in picture, below) because the chemical assault increased crosslinking between molecules. Over time, the thickening slowed the rate at which further harmful reactions took place.

**GEOSCIENCES****Reefless madness**

*Geophys. Res. Lett.* doi:10.1029/2008GL036282 (2009)

Say sayonara to coral reefs if carbon dioxide levels in the atmosphere reach 560 parts per million, roughly double their preindustrial level.

Jacob Silverman of the Carnegie Institution in Stanford, California, and his colleagues reached this conclusion on the basis of field studies that measured coral growth as a function of factors such as ocean

acidity and water temperature.

Those relationships suggest that at 560 parts per million of carbon dioxide the world's coral reefs cease to grow and start to dissolve. Unlike past studies, this one relied on measurements from actual coral reefs, rather than rates derived from studies in the lab or in large aquaria.

**NEUROSCIENCE****Spinal zap**

*Science* 323, 1578–1582 (2009)

Electrical stimulation of the spinal cord can allow rodents with symptoms of Parkinson's disease to move normally.

This discovery suggests that spinal stimulation could eventually substitute for the risky 'deep-brain stimulation' (DBS) procedure offered to some patients with severe Parkinson's disease, according to Romulo Fuentes and his colleagues at Duke University in Durham, North Carolina, who carried out the studies. DBS is an effective therapy, but highly invasive, because stimulating electrodes must be implanted deep in the brain.

The scientists say the spinal stimulation probably works through nerves that link the spinal cord to the motor cortex, where intentions to move are processed.

For a News story on this research see <http://tinyurl.com/dl2zrl>

**CONDENSED MATTER****Trolling for poles**

*Nature Phys.* doi:10.1038/nphys1227 (2009)

Might it be possible to have a north magnetic pole without a corresponding south? Despite three-quarters of a century of theoretical

investigation and occasional searches, such as unpaired poles or 'monopoles' have yet to be seen.

There is growing evidence that 'spin ices', materials containing rare-Earth ions locked into pyramidal lattices, could create something akin to monopoles as a result of the collective behaviour of their individual atoms. Ludovic Jaubert and Peter Holdsworth of the University of Lyon in France have added to that evidence. They show that experimental data for the dynamic magnetic properties of a spin ice are well interpreted in terms of the behaviour of monopoles.

## BIOLOGY

### Electric cows

*Proc. Natl Acad. Sci. USA* doi: 10.1073/pnas.0811194106 (2009)

Having previously shown statistically that cows and deer preferentially align their bodies north-south, Hynek Burda of the University of Duisburg-Essen in Germany and his colleagues now provide more evidence for these animals' magnetic sensing.

If this alignment is truly due to some sensing of the magnetic field, strong, nearby magnetic fields such as those from power lines should disrupt ruminant orientation.

Using satellite and aerial photographs, the researchers show that 1,699 cows grazing within 50 metres of overhead power lines at various European locations were randomly oriented. Field observations of 653 deer within 50 metres of power lines in the Czech Republic also revealed random orientation.

## IMMUNOLOGY

### Self help

*J. Exp. Med.* **206**, 549-559 (2009)

Dendritic cells, which present antigens to immature T cells to prime them for action against pathogens, seem to contribute to autoimmune disease. But recent findings suggest they can prevent autoimmunity as well. David Voehringer at the University of Munich in Germany and his colleagues have found that ridding mice of dendritic cells seems to cause a severe autoimmune reaction.

They created transgenic mice that selectively expressed a bacterial toxin in their dendritic cells, and had less than 10% of normal levels as a result. These mice were unable to mount an immune response to infections, but by six weeks of age their internal organs had been infiltrated by a large numbers of mature helper T cells, causing inflammation; by eight weeks this process was fatal in 40% of the mice.

## ENTOMOLOGY

### Sick bugs take drugs

*PLoS ONE* **4**, e4796 (2009)

The Very Hungry Caterpillar (pictured) in Eric Carle's children's book, which saw its fortieth anniversary this year, may have been hunting for drugs as well as food.

Woolly bear caterpillars (*Grammia incorrupta*) normally ingest small amounts of pyrrolizidine alkaloids to ward off parasites. Michael Singer of Wesleyan University in Middletown, Connecticut, and his team showed that the caterpillars ingested around twice as much of the drug when infected by maggots.

This finding suggests that self-medication is more widespread than previously thought — it has never been clearly demonstrated

E. CARLE/PUFFIN



in insects before. Previous studies by the team have shown that the taste receptors of infected woolly bears are altered to crave more alkaloids when they are infected.

## BIOMIMETICS

### Steel strong, air light

*Science* **323**, 1575-1578 (2009)

A research team led by Ray Baughman at the University of Texas at Dallas has created an artificial muscle from a ribbon of aerogel — a lightweight, sponge-like material consisting mostly of air — threaded with carbon nanotubes.

Applying a voltage across the width of the ribbon electrically charges the nanotubes; their mutual repulsion can increase the material's width by a factor of three in an instant. A voltage along the length of the ribbon triggers contraction, making the material dense and stiff. The muscles can be turned on and off 1,000 times a second. They can also be 'frozen' at any desired density, which might open up a range of applications depending on the optical and electronic properties of a specific arrangement of nanotubes.

For a News story on this research see <http://tinyurl.com/c68oum>

## JOURNAL CLUB

**Paul J. Dyson**  
Swiss Federal Institute of  
Technology, Lausanne

### A chemist highlights promising organometallic drugs.

Traditionally, the compounds we use to fight cancer come in two flavours. Inorganic drugs, such as cisplatin — a small molecule with a platinum core — are the workhorses of chemotherapy. They are generally highly toxic to cells, not particularly selective, and are accompanied by side effects ranging from vomiting to kidney damage. Larger organic drugs offer a more targeted but weaker approach. They can selectively pick off key enzymes, but may work on only a narrow range of cancers.

In the search for more effective anticancer weaponry, hybrids of inorganic and organic components — organometallic drugs — are increasingly important. Once thought of as unstable, highly toxic species, these compounds are now being developed by chemists to treat a broad range of tumours and to overcome platinum-resistant cancer cells. Like organic drugs, they have a selective mode of action and so cause fewer side effects.

The dinuclear ruthenium-arene compounds trialled by Bernhard Keppler of the University of Vienna and his colleagues are promising examples (M. G. Mendoza-Ferri *et al. J. Med. Chem.* **52**, 916-925; 2009). These highly cytotoxic compounds contain two ruthenium centres, separated by an adjustable organic linker. By tweaking the length of this chain, the researchers produced compounds that are as active as established platinum-based drugs in human tumour-cell lines.

What's more, the ruthenium drugs could kill tumour cells that were resistant to oxoplatin, a drug related to cisplatin. They work by linking DNA duplexes together, and can also bind histone proteins to DNA.

The compounds have now progressed to experiments in animals. By reducing side effects, I hope the drugs can improve the quality of life for patients undergoing chemotherapy.

Discuss this paper at <http://blogs.nature.com/nature/journalclub>



## NEWS

# Graphene gets ready for the big time

Physicists are talking about how to make practical use of a former laboratory curiosity.

## PITTSBURGH, PENNSYLVANIA

Physicists are in the grips of graphene madness. At last week's American Physical Society meeting in Pittsburgh, Pennsylvania, they packed conference rooms to hear about the atom-thick sheets of honeycombed carbon. Talks on graphene transistors, chemical sensors, electrodes, scales and frequency generators could all be heard, with participants from industry, notably IBM, in many of the sessions.

The ultra-thin carbon sheets have turned the normally staid community into "a herd of rhinos", says Andre Geim, a physicist at the University of Manchester, UK. And, he adds, "this year, I feel more like applications are what's driving the field."

Not everyone is sanguine about graphene's chances for going commercial. Graphene has several problems, notably a lack of an obvious 'band gap', a break in electron energy levels that would allow it to be easily used as a transistor, says Kenneth Shepard, an electrical engineer at Columbia University in New York. "There are a lot of problems with this stuff," he warns, fearing that starry-eyed researchers may overhype this latest material.

But others argue that graphene is much more promising than its predecessor, carbon nanotubes. Nanotubes, essentially rolls of graphene, have been difficult to control and integrate into existing electronics, says Tomás Palacios, an electrical engineer at the Massachusetts Institute of Technology in Cambridge. Graphene's uniformity and flatness make it easier to combine with current silicon technology, and many researchers who once worked on nanotubes are now focusing instead on graphene. The shift was evident at this year's meeting: there were 16 sessions on nanotubes, whereas graphene had 28.

Work on graphene — discovered by Geim and his colleagues almost 5 years ago (K. S. Novoselov *et al. Science* **306**, 666–669; 2004) — heated up quickly as researchers realized that the material's two-dimensionality caused it to show unusual quantum behaviours (see *Nature* **438**, 201–204; 2005). But graphene also

has properties that make it alluring for certain applications. Electrical charge can fly through the sheets at high velocities, up to four times those in silicon. Large thin layers of graphene would be both flexible and transparent. Graphene ribbons might act as transistors, even though bulk graphene does not. And because graphene is so thin, even the slightest brush from neighbouring atoms can alter its mechanical and electrical properties. "It has been a fascinating material," says Marcus Freitag of IBM's T. J. Watson Research Center in Yorktown Heights, New York.

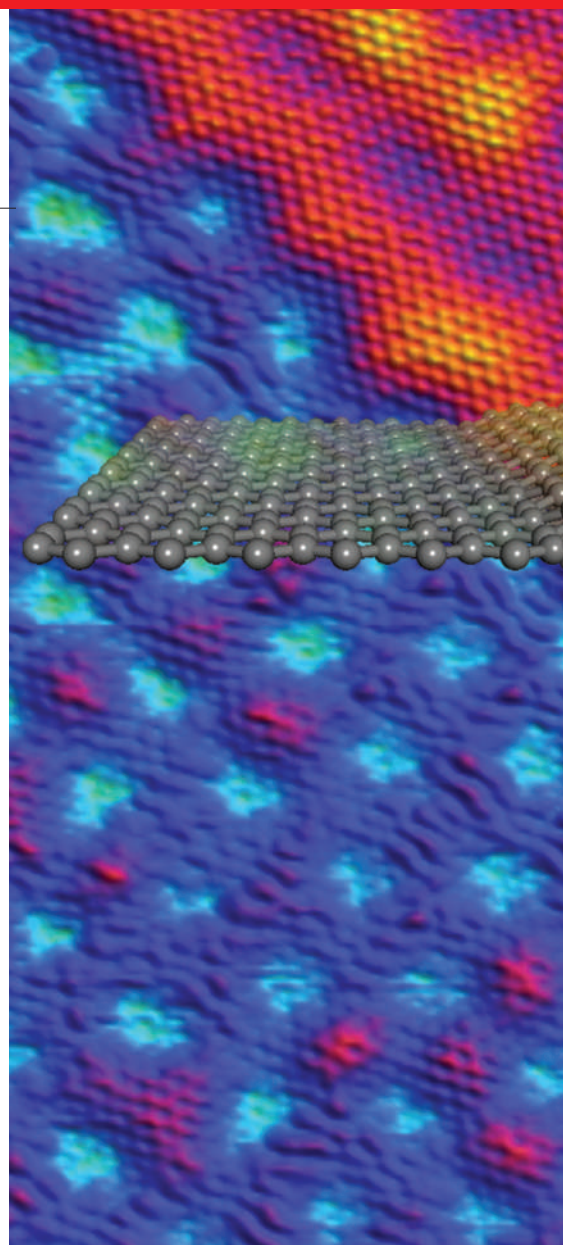
## Silicon transplant?

To turn graphene applications into reality, the material must first be synthesized in large quantities. Until now, it has often been grown on substrates of silicon carbide, a costly material that is available in only limited quantities from suppliers. But at last week's meeting, several new techniques were on display, including a way to grow graphene through chemical vapour deposition, a process widely used in the electronics industry. In one session, Byung Hee Hong of Sungkyunkwan University in

South Korea reported using the technique to grow films up to 10 centimetres in diameter — a figure he soon hopes to double. "There have been great advances in making large-scale graphene," Freitag says.

While some researchers work on making more graphene, others hunt for ways to use it. The most immediate application would be as a simple electrode. Although transparent electrodes using materials such as indium tin oxide are already commercially available, graphene's flexibility would give it an edge in solar cells and displays, says Philip Kim, a physicist at Columbia University.

Graphene also shows promise for broadband communications, in part because electrical charge can move so quickly through it. Graphene transmitters and receivers should be able to operate at frequencies on a scale of hundreds of gigahertz ( $10^9$  Hz) or even terahertz ( $10^{12}$  Hz), far better than silicon, which operates at several gigahertz, says Palacios. Higher



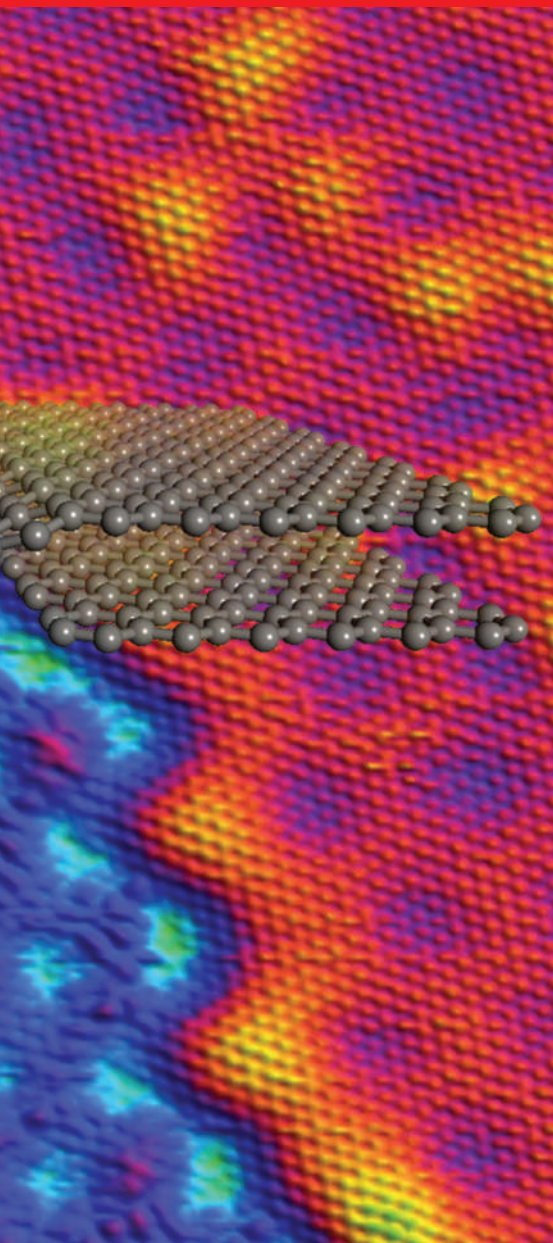
Sheet happens: graphene could have potential uses in solar cells or flexible displays.

frequencies allow for more bandwidth, and that means graphene could pave the way for broadband satellite communication. In early experiments on display at the conference, Han Wang, one of Palacios's graduate students, presented data up to one megahertz ( $10^6$  Hz), but Palacios is bullish: "We should be able to have competitive devices just a few months from now," he says.

Whether graphene can replace silicon as the basic unit of the electronics industry is another question; its lack of a band gap is a formidable problem. The most obvious solution is to cut the material into ribbons, which have discrete energy levels. But, as several groups showed in Pittsburgh, cutting the sheets creates a jagged edge of dangling chemical bonds that can pick up unwanted contaminants. Xinran Wang of Stanford University in California reported some success in using ammonia and other compounds to dope the edges of the graphene

A. WEE, NATL UNIV. SINGAPORE/H. HUANG ET AL. ACS NANO **2**, 2513–2518 (2008)





ribbons, allowing them to carry charge more easily. Even then, the scattering of electrons from the ribbon's ragged edges greatly reduces its performance. Based on these kinds of findings, Shepard believes that making these devices work will be extremely difficult. "Nothing's going to supplant silicon, not in my lifetime," he says.

Ultimately, it may be too early to tell just what graphene will — or won't — be able to do. As groups presented models and raw data from their early graphene gadgets, it became clear that many are still grappling with the latest addition to the pantheon of carbon materials. At the end of one talk, Andrea Carlo Ferrari of the University of Cambridge, UK, flashed a few slides onto the screen. Apparently, oxidizing graphene causes it to glow under infrared laser light, Ferrari told the crowd. The data are fresh, and the implications still unclear. "Will this lead somewhere?" Ferrari said afterwards with a shrug. "We don't know."

Geoff Brumfiel

## UK funding ban sparks protests

British scientists are campaigning against a plan to bar hundreds of unsuccessful grant applicants from making funding bids in the following year.

The rule, announced by the government's Engineering and Physical Sciences Research Council (EPSRC) on 12 March, aims to reduce the pressure on an overloaded system that currently peer reviews all grant applications.

But by 24 March, more than 1,200 protesters had signed an online petition (<http://tinyurl.com/cvyexx>) demanding that the policy be repealed. "The feeling in the community is that it is draconian and deeply unfair," says Philip Moriarty, a physicist at the University of Nottingham, UK. He and other scientists contacted by *Nature* say they will refuse to review their colleagues' work under such a system.

Science-funding experts think that the strategy is unique among UK, US and European funding bodies. "We could not do it in the United States. It would be very contentious," says Antonio Scarpa, director of the Center for Scientific Review at the National Institutes of Health in Bethesda, Maryland. Frank Wissing, life-sciences programme director at Germany's science funding agency, the DFG, adds that its committees have never discussed a ban on unsuccessful applicants.

The EPSRC says that scientists will not be allowed to apply for research funding for 12 months if, in the past 2 years, they have had three or more proposals ranked in the bottom half of a funding prioritization list, and also have less than 25% of all their proposals funded in that time.

The funding council says it expects 200–250 researchers will be excluded, accounting for 5% of applicants but 10% of the total number of applications submitted to the council. Those researchers are "producing a disproportionate load on the peer-review system," says Bill Wakeham, vice-chancellor at the University of Southampton, UK, and a member of the EPSRC council. Excluded scientists will have to undergo a mentoring programme to help improve their success rates before being allowed to submit grants again.

The exclusions will begin from 1 June, and the EPSRC will be contacting the first

wave of scientists affected in April, says David Delpy, EPSRC chief executive. "We are a little uncomfortable with something that is applied retrospectively," he says. "But we can't wait another two years to implement it, with success rates falling as they are." Although that rate has hovered around 30% since 2004, it has dropped to just 24% in the past year.

Low success rates make applicants more cautious about spending time preparing ambitious proposals, says Delpy, and a flood of safer proposals could crowd out higher-risk but potentially ground-breaking ideas.

Chemists are most likely to be affected by the policy, says David Reid, head of marketing and communications for the EPSRC, because they tend to submit larger numbers of smaller, short-term proposals compared with other subject areas. Some funding areas with a focus on chemistry have seen success rates fall as low as 15%.

"It is the chemists who are mostly complaining, and it is the chemists who produce most of the applications that fail," says Wakeham.

Tom Welton, head of chemistry at Imperial College London, echoed the feeling of many chemists contacted by *Nature*, calling the move a "knee-jerk bureaucratic response". "We are

appalled by the lack of consultation," adds Joe Sweeney, an organic chemist at the University of Reading, UK.

Reid concedes that the EPSRC did not consult widely on the specifics

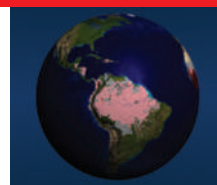
of the policy. But he argues that a 2007 consultation by Research Councils UK, an umbrella group for the country's research funding councils, had found that some academics supported the idea of targeted disincentives to improve success rates.

Delpy says that other options considered by the council, but rejected, included charging for submissions; applying institutional quotas; or penalizing universities by making doctoral training grants proportional to their success rates.

Along with the exclusion policy, which will be reviewed in a year's time, the EPSRC will also refuse uninvited resubmissions of failed proposals, bringing it in line with other UK research councils.

Richard Van Noorden

See Editorial, page 385.



**MAPPING MALARIA**  
Most detailed atlas of  
disease risk unveiled.  
[www.nature.com/news](http://www.nature.com/news)

MALARIA ATLAS PROJECT/NASA

# Ship decision sinks UK marine science

Citing increased costs due to the weakness of the pound against the euro, the British government has pulled the plug on a £55-million (US\$80-million) order to deliver a replacement vessel for its ageing research ship, the RRS *Discovery*, by 2011. The move, which could see the ship pushed further down the line, is a major setback to UK marine science, say researchers. Re-ordering the ship will add years of delay just as expeditions are being postponed or cancelled because the 47-year-old *Discovery* frequently breaks down.

The *Discovery* entered service in 1962, and is one of two UK ocean-going research ships widely available for marine science, both of which are operated by the Natural Environment Research Council (NERC). The other is the RRS *James Cook*, which made her maiden voyage in 2007 but has been plagued by bubbles generated by her hull (see "Troubles bubble up"). The British Antarctic Survey has another ship, the RRS *James Clark Ross*, but this is available only 60 days a year for research outside the survey.

NERC and the government had agreed to split the costs of the replacement for the *Discovery*. She was intended to be built at a Spanish shipyard, the only bidder that responded to the agency's call for tender two years ago, in a then-saturated ship-building market. But NERC officials learned on



S&G AND BARRATTS/PA

**Problems of age:** the RRS *Discovery* has been in operation since 1962, and suffers frequent breakdowns.

11 March that the Department for Innovation, Universities and Skills (DIUS) had torpedoed the current order, citing the unfavourable exchange rate between the pound and the euro. Matthew Barker, a spokesman for the department, declined to comment further, noting only that an official memo on the topic promised that "NERC and DIUS are committed to

procuring a replacement for *Discovery* as soon as possible subject to affordability".

NERC will discuss its options at a board meeting on 1 April, but the cancellation means that, at best, the procurement process will probably need to start from scratch. The project as originally envisaged might also unravel, with alternative options being a smaller ship or the conversion of a second-hand vessel. "The bottom line is that there will be a

delay, and that delay will be fairly serious," says George Wolff, an oceanographer at the University of Liverpool, UK, and senior representative on the RRS *Discovery* Replacement User Group.

Ship time is already so stretched that scientists getting an expedition approved now will have to wait until after 2012 to set sail, as the ships are fully booked until then. And even this scheduling is based on the assumption that the *Discovery* will hang together. Researchers familiar with the vessel aren't so optimistic. "They are keeping her going, but she's liable to fall apart at any moment," says Wolff.

"There has been a desperate need to get this new ship," adds Malcolm Woodward of the Plymouth Marine Laboratory, who represents UK marine-chemist users on the *Discovery* replacement group. Woodward knows the problems first hand: later this month, he will try for the third time to set sail on an expedition to study the interaction between the ocean surface and the lower atmosphere. Two previous

## Troubles bubble up

On 5 February 2007, Britain's Princess Anne smashed a champagne bottle on the hull of the £40-million (US\$58-million) RRS *James Cook* (pictured), christening the newly built vessel. Researchers have since discovered that the ship's hull is associated with bubbles other than of champagne, which are far less cause for celebration.

The ship has a bulbous bow and a flat hull, similar to those of oil tankers, that make for better seafaring performance. But on the *James Cook* the design also creates a sweep of microbubbles beneath the hull, which had not been foreseen. The problem "knackers all the acoustics instruments",

including the ship's multibeam swath bathymetry for creating depth and other maps, says one scientist.

"There is an issue of degraded performance," concedes Natural Environment Research Council (NERC) spokesman Edward Hill, director of the National Oceanography Centre in Southampton, UK. He says that performance is better or worse depending on such factors as wave conditions and cruising speeds. NERC has created working groups to see if there is a way around the problem, he adds, but so



far "the outcome has been inconclusive".

"The bubble sweepdown problem makes the ship non-ideal," says George Wolff, an oceanographer at the University of Liverpool, UK. "One of the most important considerations we have had with the design of the new RRS *Discovery* is to ensure that this does not happen again." **D.B.**

P. BARRITT/ALAMY



**TRAPPED UNDER ICE**

Risky submarine mission maps the underside of an Antarctic glacier.

[www.nature.com/news](http://www.nature.com/news)

BAS

attempts over the past three years had to be cancelled when the *Discovery* broke down.

The new vessel would not only have replaced the *Discovery* but would also have incorporated twenty-first-century technology, such as mini-thrusters, which would allow the ship to remain in a precise position for sampling a water column. And she would have featured multibeam swath bathymetry, an acoustic technique used to produce depth and other maps, together with instruments central to geophysical and other research

that are lacking on the *Discovery*.

Speaking on behalf of NERC, Edward Hill, director of the National Oceanography Centre in Southampton, puts a brave face on the government's decision. "It's a disappointment," he says, although he notes that it's a postponement, not an outright cancellation.

To make the high costs of research ships more palatable to governments, scientists in the 1990s touted the idea of sharing the cost of building new research vessels by making them European instead of national. In 1996, Britain, France and

Germany agreed to barter ship time on each other's vessels as a first step towards this goal. The Netherlands, Norway and Spain have since joined them. Ironically, the scheme has proved so successful that the goal of a pan-European fleet has all but been abandoned. But bartering time on other countries' ships demands having something to barter with, notes Wolff. "The UK is rapidly running out of barter credit, and this decision to postpone its new ship has diminished it even further."

**Declan Butler**

## Canadian science minister under fire

It's been a rough month for Canada's minister of science and technology. Gary Goodyear, who was appointed to the new position in October 2008, has been roundly criticized in the media for an outburst during a meeting with a university teachers' group and for his comments on evolution.

Some Canadian researchers say the criticism is unreasonable, but others say it suggests that Goodyear, a chiropractor by training, is not in tune with the community whose portfolio he oversees.

In early March, the Toronto-based *Globe and Mail* reported that Goodyear had stormed out of a meeting with two officials from the Canadian Association of University Teachers (CAUT), a lobby group representing 65,000 academic staff at 121 universities and colleges. The spat erupted over the group's concerns about the 2009 budget, which gave investments in university infrastructure priority over operational research funding (see *Nature* 457, 646; 2009). Some argue that the government is emphasizing the commercialization of science at the expense of basic research.

Then on 17 March, Goodyear's scientific credentials were called into question by the

headline "Minister won't confirm belief in evolution". According to the *Globe and Mail*, when asked whether he believed in evolution, Goodyear responded, "I'm not going to answer that question. I am a Christian, and I don't think anybody asking a question about my religion is appropriate". During a later interview on television, Goodyear declared that he does accept evolution and added that he didn't answer the question because "it wasn't relevant to the portfolio" and that the "focus should be on the economy and creating jobs".

Some scientists say the flap over evolution is unfortunate. Many couldn't care less about his personal beliefs. "As a professional

he would never allow his religion to come into play in work. We all expect that level of professionalism," says Ryan MacKay, scientific director of the National High Field Nuclear Magnetic Resonance Centre in Edmonton, Alberta.

But others were disturbed by the words Goodyear used in the television interview. "We



Gary Goodyear had a surprising take on evolution.

are evolving every year, every decade," he said. "That's a fact, whether it's to the intensity of the sun ... or to the effects of walking on concrete. Of course, we are evolving to our environment."

"Canadians can rightly question the technical competence of elected officials," says Robert Wolkow, a nanoscience expert and physicist at the University of Alberta in Edmonton. "Just as we would expect an agriculture minister to know a heifer from a hog, it's reasonable for a science minister to know established scientific basics," he said. Goodyear was unavailable for comment for this article.

Goodyear reports to parliament on the activities of the Natural Sciences and Engineering Research Council (NSERC) and the Canadian Institutes of Health Research (CIHR) and is responsible for their budgets, although

grants are determined by peer review. He is also in charge of implementing the Science and Technology Strategy, which was released in May 2007.

At this point no one seems interested in calling for Goodyear's resignation. James Turk, executive director of the CAUT, says that the question is not whether Goodyear should be replaced but whether the government will commit to investing in research. "We need the government to match what the Obama government is doing for research," he argues.

The 2009 federal budget has been widely criticized by the research community. Much of the worry has centred on its effects on the three granting councils — the NSERC, CIHR and the Social Sciences and Humanities Research Council — which were told to scale back their collective budget by Can\$147.9 million (US\$120 million) over three years.

"It's a significant concern for those of us who believe that curiosity-driven research will, over time, lead to major breakthroughs," says Dick Peltier, a physicist and the director of the Centre for Global Change Science at the University of Toronto. "It's a short-term mindset that is not in accord with the historical evolution of ideas, and it is a terrible mistake."

**Hannah Hoag**

C. WATTIE/REUTERS



# More pain studies needed

Scientists struggle to fund work on animal pain and distress.

More research is needed on how to determine when a laboratory animal is in pain and how best to alleviate that pain, says a report from the US National Research Council. That conclusion may sound familiar; a report last year called for more research into laboratory animal distress, and a 1992 study said that more work was needed on pain and distress.

However, little has been done because the root problem — lack of money for such studies — has remained for decades. “Unless or until someone comes forward with the funding, it is not going to happen,” says Gerald Gebhart, a pain researcher at the University of Pittsburgh in Pennsylvania and chairman of the committee that put together the new report.

Research into animal pain tends to fall between the purviews of existing funding agencies, the biggest of which typically have human health as their goal. “All of what we know today about pain in animals — well, 95% of it — comes from research on animals as models for human

pain,” says Andrew Rowan, a vice-president at the Humane Society of the United States, in Washington DC. For example, information on which pain medication regime is ideal for a mouse that has just had abdominal surgery might be gleaned from an experiment focused on what pain regime is ideal for humans.

In the United States, grants solely to study animal pain are rare and small. The funding situation is better across the Atlantic

Ocean. “In Europe there is a lot of interest,” says Vera Rogiers, head of the European Consensus Platform on 3R Alternatives, a non-profit organization based in Brussels. A proposal released late last year would tighten up a European

Commission directive on the care of laboratory animals (see ‘Europe to revise animal-testing rules’), which, among other things, would virtually ban studies on great apes, consider the pain of invertebrate species as well as vertebrates and enshrine the desirability of the three Rs: refinement, reduction and replacement.

**“Pain research hasn’t worked well because our animal models are inappropriate.”**



Paul Flecknell, an author of the US report at the University of Newcastle, UK, has had grants from the Wellcome Trust in London, and the UK Biotechnology and Biological Sciences Research Council, among others, to maintain his studies on behavioural clues to animal pain and distress (A. L. Dickinson, M. C. Leach and P. A. Flecknell *Lab. Anim.* 43, 11–16; 2009). Still, he says, money is in “very short supply” for studies in which the primary goal is to improve practices in the use of lab animals.

J. McDONALD/CORBIS

# Bills target biosimilar drugs

Makers of biological drugs and their would-be generic competitors drew new battle lines on Capitol Hill last week, in a long-simmering debate over when and how generic versions of these large, complex drugs will enter the US market. Two California Democrats have introduced duelling bills in the House of Representatives; which one emerges as the winner — or what compromise is struck between the two — will indicate how much power and influence the industry lobby has.

In his proposed budget for 2010, President Barack Obama supports allowing the US Food and Drug Administration (FDA) to approve generic versions of biologicals, also known as biosimilars. As he seeks to cut soaring health costs, the high prices of biological drugs present an attractive target. Last year, the Congressional Budget Office estimated that a Senate bill creating a regulatory pathway for biosimilars would cut spending on prescription drugs in the United States by \$25 billion between 2009 and 2018.



Anna Eshoo and Henry Waxman go head-to-head over drug regulation.

Other countries have set a precedent for such approval. The European Union opened up a regulatory pathway for biosimilars in 2005, and since then its European Medicines Agency has approved 13 such drugs. A

number of the European guidelines have been adopted in Australia, and Japan this month issued its own guideline for the regulation of biosimilars.

In the United States, Anna Eshoo, the Democrat who represents California’s biotech-heavy Silicon Valley in Congress, has introduced the Pathway for Biosimilars Act with leading co-sponsor Joe Barton (Republican, Texas). And on 11 March, Henry Waxman, the energy and commerce committee chairman from Los Angeles, introduced the Promoting Innovation and Access to Life-Saving Medicines Act, co-sponsored by Nathan Deal (Republican, Georgia) and others.

Both groups agree that a law allowing the FDA to approve biosimilars should preserve patient safety while opening up the market to healthy competition — all without squelching incentives to innovation. But they agree on almost nothing in the legislative details of how to achieve this. “The biotechnology industry has endorsed the Eshoo bill, and the generic industry has endorsed the Waxman bill. It doesn’t get more extreme than that,” says Mark Schoenebaum, a biotechnology analyst in

S.J. FERRELL/CONGRESSIONAL QUARTERLY/NEWS.COM

**OIL-SPILL ANNIVERSARY**

Twenty years on, the lingering legacy of the Exxon Valdez.

[www.nature.com/news](http://www.nature.com/news)

BETTMANN/CORBIS

## Europe to revise animal-testing rules

Animal pain is one of a host of issues facing legislators, as a major revision of rules governing animal experiments makes its way through the European Parliament.

The proposals — a wholesale update of the decades-old current rules — were unveiled last year by the European Commission. The parliament's agriculture committee is scheduled to vote on them on 31 March. If the legislation can clear this committee quickly, it may be voted on by the

whole parliament before the European elections in June.

Under the legislation, European Union states would have to "ensure the improvement of breeding, accommodation and care, and of methods used in procedures, eliminating or reducing to the minimum any possible pain, suffering, distress or lasting harm to the animals".

Another proposed rule suggests that all experiments should be categorized on the

basis of the pain and harm they may cause, and that lab animals should generally only be re-used in experiments classed as 'up to mild'. The exact definitions of these categories are not yet set.

Some scientists have warned that some of the rules, such as those on non-human primates, and an expected increase in bureaucracy, could make it harder to perform research on animals, as well as driving up the cost of research. ■

**Daniel Cressey**

Some of his colleagues in other countries have cobbled together funding from veterinary charities and pharmaceutical companies for work on dogs and cats, and from agricultural agencies interested in poultry production for work on birds.

Animal-welfare organizations tend not to fund such research, as they focus on developing alternatives to animal research instead of refining current laboratory practices. And virtually no one funds research on pain and

distress in the non-mammal vertebrates, such as fish, or the invertebrate worms commonly used in neurology studies.

Supporting such research might one day have pay-offs for people. "Some of us in the pain field think that if one thinks that pain research hasn't worked particularly well — there are no new drugs in the clinic, really — that our animal models are inappropriate," says Jeffrey Mogil, a pain specialist at McGill University in Montreal, Canada.

Researchers who study human pain use animal models measuring how sensitive an animal is to touch after a painful stimulus. This 'mechanical allodynia' measures tenderness such as the pain evoked when a bruise is touched. But it does not measure 'spontaneous pain', such as the dull throb of a bruise that is not being touched. A better model of that in animals, says Mogil, could mean better treatments one day for pain in humans. ■

**Emma Marris**

New York for Deutsche Bank.

The starkest difference between the two bills — and probably the point of most heated contention in the upcoming congressional debate — is the period of market exclusivity they grant. Under Waxman, an innovator company would be guaranteed at most six years of competition-free market access. Under Eshoo, it would receive at least 12 years, and up to 14.5 years. With annual sales of leading biologicals such as Amgen's Epogen — set to come off its patents in the United States beginning in 2012 (see table) —

at more than \$2.4 billion, the stakes in what number is finally settled on are huge.

Significantly, Waxman has given ground from earlier versions of his bill, which guaranteed innovators no period of market exclusivity. Eshoo's bill, by contrast, remains essentially the same as it was in the last Congress, when a bipartisan group of leading senators also endorsed a 12-year exclusivity period. This year other senators are also expected to introduce a bill that, like Waxman's, endorses a 6-year period.

The Eshoo bill also proceeds more cautiously on other key issues, such as whether a biosimilar is interchangeable with the innovator drug, which could allow pharmacists and health plans to switch patients without a doctor's go-ahead. Before the FDA could allow this for a given biosimilar, Eshoo's bill would require the FDA to issue a 'guidance' document — an often-lengthy and public process — establishing the scientific grounds for allowing the change. Waxman's bill has no such requirement.

"Cost saving is obviously a goal here," says Sandi Dennis, the deputy general counsel for health care at the US Biotechnology

Industry Organization. "But you can't make decisions purely on cost savings without first assuring that the appropriate data are there for a new product to assure patients' safety. The science has to precede the economics."

Generics makers argue that such requirements are stalling tactics. "In these tough economic times, the Eshoo-Barton bill is a dead end for consumers," says Kathleen Jaeger, president and chief executive of the Generic Pharmaceutical Association in Arlington, Virginia. "It's just riddled with barriers. Consumers are not going to get timely access to affordable medicines."

Both bills give discretion to the FDA to work out approval requirements, such as whether clinical trials are needed to win access to the market. Brand-name makers argue that trials are crucial because biosimilars, unlike generic copies of small-molecule drugs, are not exact copies of innovator biologicals and could thus cause unexpected safety problems, such as immunological reactions. Generics makers retort that the FDA has sufficient scientific expertise to judge when trials are needed — and shouldn't be required to impose them. ■

**Meredith Wadman**

### KEY PATENTS DUE TO EXPIRE

Expiry	Brand name	Generic name	Maker
2010	Cerezyme	Imiglucerase	Genzyme
2012	Epogen	Epoetin alpha	Amgen
2012	Enbrel	Etanercept	Amgen
2012	Avonex	Interferon $\beta$ 1a	Biogen
2013	Neupogen	Filgrastim	Amgen
2014	Rituxan	Rituximab	Genentech
2015	Neulasta	Pegfilgrastim	Amgen
2015	Fabrazyme	Agalsidase beta	Genzyme

SOURCE: M. SCHOENENBAUM



# NOAA chief ready to tackle climate

A little more than a decade ago, ecologist Jane Lubchenco set out to train scientists how to communicate with non-scientific audiences and thus create better national environmental policy. The Aldo Leopold Leadership Program has since become one of the most respected training programmes for senior scientists.

Now, after guiding more than 100 researchers through the treacherous waters of congressional relations, Lubchenco is putting her own advice into practice as the new head of the National Oceanic and Atmospheric Administration (NOAA). Confirmed by the US Senate on 19 March, along with John Holdren as the director of the Office of Science and Technology Policy, Lubchenco takes the helm of a US\$4.4-billion agency charged with overseeing research and monitoring in fields from marine mammal populations to climate change.

Lubchenco, the first woman to head NOAA, is going through a series of detailed briefings to get her up to speed on the agency's many doings. But she is already speaking in generalities about her priorities. "NOAA has a very distinguished track record of science," she says. "We're going to build on that track record and use the science that we produce to serve the nation." Her wish list is nothing if not ambitious: she says she wants to "solve the overfishing problem", put the agency's earth-monitoring satellites programme back on track, establish a National Climate Service to provide climate-related data to users, and "protect and restore ocean ecosystems".

"She is one of the best scientists to give advice to Obama, arguing with the Larry Summerses of the administration and their economic theories," says Stephen Schneider, a climatologist who directs Stanford University's Center for Environmental Science and Policy in California, and who has been a long-standing adviser to the Leopold programme.

A marine tidal specialist at Oregon State University in Corvallis, Lubchenco has for decades shared research, training and a lab with her biologist husband, Bruce Menge. "They are very different," Schneider says. "Jane cruises the Black Sea now with archbishops discussing stewardship of the planet. Bruce prefers to head to the reefs to count species."

Lubchenco has, in fact, long juggled research, policy and activism, and her long list of accomplishments includes past presidencies of the American Association for the Advancement of



A. GEA/REUTERS

Jane Lubchenco has been confirmed as head of the National Oceanic and Atmospheric Administration.

Science and the Ecological Society of America. She has not, however, ever run a major governmental agency or facility.

Karina Nielsen, of Sonoma State University in Rohnert Park, California, was inspired to become a marine ecologist when she read quotes attributed to Lubchenco in an undergraduate biology textbook. Eventually, she received her doctorate in the Lubchenco/Menge lab. "They have an incredible vision of letting students be creative," she says. "You make mistakes, but you learn how to do excellent science. Encouraging creativity is the hallmark of their lab." For example, Nielsen's doctoral thesis challenged a widely held view that intertidal basins blossomed with life based mainly on the upwelling of basic nutrients.

"Jane is a pioneer woman in science — a role model combining quality research and leadership," says Lisa Curran, an ecologist at Yale University and a Leopold fellow. The Leopold training helped in fighting the timber industry to establish the Center for Tropical Forest Science site in Borneo, Curran says. "I became much more confident speaking", and was able to present publicly as more than just a scientist "with dirty dresses and leech bites".

Now, about 20 Leopold fellows are trained in two week-long sessions, one of which is in Washington DC and includes visits to Congress. From the start, Lubchenco envisioned it as a programme to stretch across environmental

disciplines and push boundaries. "I was included in the first cohort, I think, because I'm an ecologist who studies infectious diseases," says Andrew Dobson of Princeton University in New Jersey. "It taught me to be really organized, to get my ideas across clearly."

The Leopold training was even helpful for seasoned veterans of political skirmishes such as Barry Noon, a modeller at Colorado State University in Fort Collins who served as a main governmental witness when the timber industry sought to cut old-growth forest in the northwestern United States, including habitat of the northern spotted owl. "I was always disappointed in my testimony, feeling I could have done better," he says. Years later, when he did the Leopold fellowship, he realized how such training could have helped.

Lubchenco will have plenty of chances to use her well-honed communications skills. On the question of overfishing, for example, industry groups have been wary of her pro-environment stance in the past. As administrator, she acknowledges the competing priorities. "The health of fish stocks is directly related to the health of many coastal communities, and we need to find a way forward that balances all of the different concerns," she says. "It will be a very tough challenge, but it is doable and we must do it."

Rex Dalton with reporting by Alexandra Witze.

For a Q&A with Jane Lubchenco, see <http://tinyurl.com/cjuogy>.

**"NOAA is like city states where people don't talk to each other."**

## Funder moves to quell Turkish censorship row

The publisher of Turkey's most popular science magazine, *Bilim ve Teknik* (*Science and Technology*), has resolved a headline-hitting dispute over the magazine's apparent censorship of an article about Charles Darwin (see *Nature* 458, 259; 2009). The row was politically charged because the magazine is published by TÜBİTAK, the national science-funding agency.

The agency's vice-president, Ömer Cebeci, withdrew a cover and 16-page feature devoted to Darwin just before the March issue went to press. He claimed that the editor, Çiğdem Atakuman, had secretly changed previously agreed content, and he told her that she would be transferred to another department. TÜBİTAK has now issued a statement confirming its commitment to science and scientific literacy in the country, where many people do not believe in evolution. Atakuman will remain as editor of the magazine.

## Marine sequencing project embarks on another cruise

Scientists from the J. Craig Venter Institute have reboarded the research sloop *Sorcerer II* to sample microbes from the Baltic, the Mediterranean and the Black seas. The microbes will be frozen and sent back for DNA sequencing at the institute's labs in Rockville, Maryland, and San Diego, California.

Craig Venter and his colleagues have already circled the globe to collect open-ocean microbes (*Nature* 446, 240–241; 2007). The new voyage, which commenced on 19 March, will “give us substantially more diversity than we find in the open oceans of the world”, says Venter.

Flanked by dense human populations, the three seas are likely to paint a portrait of human effects on microbial populations, he says. The project is expected to cost around US\$10 million and will be funded by the Beyster Family Foundation in San Diego, with additional grants from-



Setting sail: *Sorcerer II* is heading out to sea again.

J. CRAIG VENTER INST.

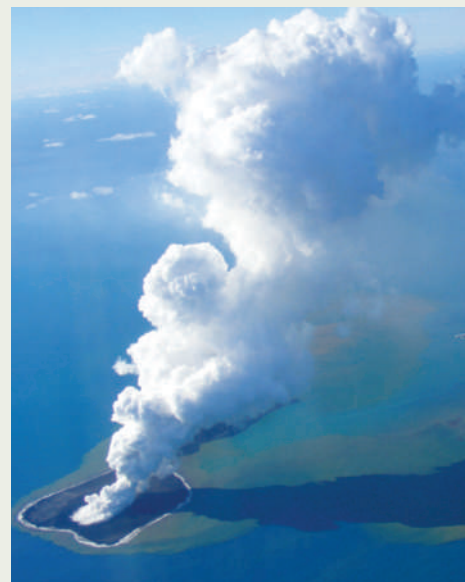
## Flames burn higher in Pacific ‘ring of fire’

A volcano in the Tonga archipelago has destroyed all plant and animal life on the island of Hunga Ha’apai, according to a survey of the area.

The volcano, which began erupting on 16 March, is sited 63 kilometres northwest of the Tongan capital, Nuku’alofa. Concern grew on 20 March when a magnitude-7.9 earthquake some 210 kilometres southeast of Nuku’alofa led to tsunami warnings for all residents within 1,000 kilometres of the epicentre.

Although the tsunami warning was later cancelled, scientists have warned that the quake could lead to more lava flowing from the volcano’s two vents: one situated on the island, the other just offshore.

Meanwhile, on the other side of the Pacific ‘ring of fire’ — the chain of volcanoes that skirts the Pacific Ocean — Mount Redoubt in Alaska erupted on 22 March, throwing huge plumes of ash into the sky.



T. FOTU/MATANGI TONGA/AFP PHOTO/GETTY IMAGES

anonymous donors and Life Technologies, the biotechnology company formed from the merger of Invitrogen and Applied Biosciences.

## US medical institute to set up African research centre

The Howard Hughes Medical Institute (HHMI) in Chevy Chase, Maryland, last week announced the setting-up of its first research laboratory outside the United States. The KwaZulu-Natal Research Institute for Tuberculosis and HIV in Durban, South Africa, will operate in partnership with the University of KwaZulu-Natal and will focus on the increasing incidence of tuberculosis in those infected with HIV.

One in ten individuals worldwide is infected with *Mycobacterium tuberculosis*, the bacterium that causes the disease, but it does not usually cause full-blown disease. However, *M. tuberculosis* can be devastating in individuals whose immune systems are weakened by HIV. South Africa accounts for almost one-fifth of the global HIV disease burden, and tuberculosis, including multi-drug-resistant and extensively drug-resistant strains, is a major problem there. The HHMI will invest US\$60 million in the centre over the next decade, and will bring in the cream of its international scientists to help train local researchers.

## US panel to guide treatment comparison

The 15-person committee that will oversee the spending of US\$1.1 billion on ‘comparative effectiveness’ research was

named by the US Department of Health and Human Services on 19 March. Elizabeth Nabel, director of the National Heart, Lung and Blood Institute in Bethesda, Maryland, will represent the National Institutes of Health, which is receiving \$400 million of the money.

Over the next 18 months, the initiative will support research by federal agencies on the relative strengths and weaknesses of various medical interventions, ideally providing physicians and patients with the information needed to drive cost savings in the world’s most expensive health-care system (see *Nature* 458, 7; 2009). It was funded as part of the \$787-billion economic stimulus package signed into law in February.

## Science head of energy department is named

Steven Koonin, the former provost of the California Institute of Technology in Pasadena and current chief scientist of global energy company BP, has been selected by US President Barack Obama to serve as undersecretary of science at the Department of Energy. Koonin would replace Raymond Orbach in overseeing the department’s \$4.7-billion science portfolio, including running the national laboratory system. The position requires Senate confirmation.

The appointment had been rumoured in Washington for some time; Koonin is a long-time colleague of Steven Chu, the new energy secretary. Both played a major part in setting up a \$500-million industry–university alliance between BP, the University of California at Berkeley and the University of Illinois at Urbana-Champaign.





Fish farming demonstration near Orlando, Florida.

# FUTURE FISH

The only way to meet the increasing demand for fish is through aquaculture. **Daniel Cressey** explores the challenges for fish farmers and what it means for dinner plates in 2030.

Sitting in an unremarkable family restaurant a short drive from his institute in Stirling, UK, Randolph Richards scans the menu's seafood offerings with an expert eye.

"The salmon is probably farmed in Orkney," he says, referring to an archipelago north of mainland Scotland. "The sea-bream — probably grown in Greece." Thus, the head of the University of Stirling's Institute of Aquaculture reveals a secret that most diners are blissfully unaware of: farmed fish are everywhere. Roughly every other morsel of fish passing through human lips was raised under human supervision.

Right now, more than 50 million tonnes of glassy-eyed livestock are corralled in underwater cages and tanks, crowded, fed, sometimes dosed with antibiotics and ultimately culled and shipped up to half-way around the world to meet the ever-growing demand for fish. It's the world's fastest growing food sector.

From decadent fatty tuna belly savoured in the most expensive sushi joints to a work-horse frying fillet such as tilapia, farmed fish are becoming the norm rather than the exception, even if customers are unaware of the subtle change on their plates.

Back in the 1970s only 6% of the world's

food fish came from aquaculture. By 2006, that proportion had risen to almost half, according to the biennial State of World Fisheries and Aquaculture report released last month by the Food and Agriculture Organization of the United Nations (FAO)<sup>1</sup>.

And to keep up with world appetites, the fish-farming industry will have to continue this trend. The reason is simple, says Rohana Subasinghe, a senior fishery resources officer at the FAO. "We are not going to get adequate fish from the sea in the coming years."

Current projections suggest that by 2030 the world's population will have exceeded 8 billion people. Maintaining today's consumption rates, of around 17 kilograms per person per year would require an extra 29 million tonnes of fish. Meanwhile, around half of all fish stocks have been deemed "fully exploited" by the FAO, with those deemed "overexploited, depleted or recovering" now around 30%.

As a result, the fishing industry is casting about for anything even vaguely palatable. Although some consumers profess an aversion to farmed fish, claiming that they're bland, uninteresting or unnatural, many would choose a boring farm-raised salmon over fresh-caught pelagic delicacies such as jellyfish and krill.

**"Even though the salmon look like fish they are almost more similar to pigs."**

— Carlos Duarte

"There's only so much wild stock out there," says Michael Rubino, manager of the aquaculture programme run by the National Oceanic and Atmospheric Administration in Silver Spring, Maryland. Most of the demand will have to be met by aquaculture, but what those future meals will look like depends very much on how much science can contribute to the trade. Predators, such as salmon and cod, are popular and command high prices at market. But their carnivorous diets rely on the same fish stocks that are under threat around the world. Tilapia — omnivorous cichlids — are fairly simple to raise. As one of the fastest growing aquaculture products, fish such as these may represent the future.

## From paddies to pools

Aquaculture has been practised in China for many thousands of years. Carp were left to grow in ponds and rice paddies and later harvested. This passive, pastoral method, with little or no attempt to actively nurture the animals, is still practised widely. China produces 67% of the world's farmed seafood, much of it carp, and much of it through these generally low-tech methods.

Another short drive from his institute, and Richards is showing off a landmark piece of fish-farming's technological lurch forward. The Howietoun fish farm, built in the 1800s by a Victorian landowner, is still a commercial farm

VISIONSOFAMERICA/J. SOHM/GETTY

K. SCHAFER/ZEEFA/CORBIS

today. Brick-lined ponds, set on a gentle slope, teem with brown trout waiting for a few scoops of food and destined to stock lakes for sport fishing. When the fish in the uphill ponds reach a predetermined size, they are shuttled to the next via connecting channels.

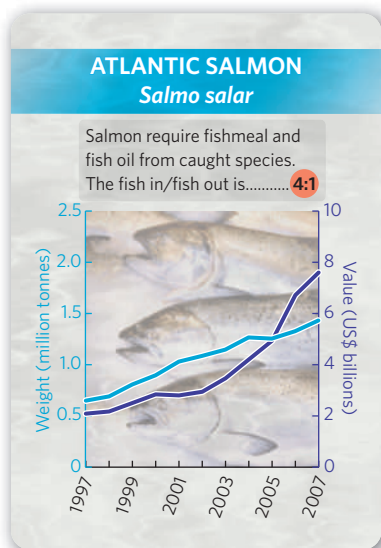
The modern descendants of Howietoun have been attempting to drive up yields, along with profits, through the help of institutions such as the one Richards heads. Starting with veterinary science and progressively incorporating more and more disciplines, ranging from ecology to genetics, these research centres have helped to dramatically increase yields, but in doing so, they have created new problems.

The quaint, idyllic farm at Howietoun — where interference from otters and herons represent the extent of environmental conflict — is a far cry from modern industrial farms, which can involve miles of cages off the coast or huge collections of tanks on land.

The most dramatic change in aquaculture has perhaps been the explosive growth in shrimp farming in southeast Asia. Encouraged by high demand and foreign investment, more than 109,000 hectares of mangrove swamp have been cleared for shrimp farming in the Philippines, for instance, since farming started in earnest in the 1970s. That's roughly two-thirds of the nation's area for these unique watersheds.

Furthermore, the nutrients added to many fish farms or produced as waste, particularly nitrogen and phosphorous, can trigger massive plant and algal blooms in surrounding waters. As the blooms die off, the bacteria taking part in the bonanza of decomposition suck the oxygen out of the system leaving it dead.

These environmental effects have caused many to oppose the idea of aquaculture, which the FAO says threatens future development in many regions. In response, farmers have been asking researchers to advise them on how much waste the surrounding sea can take. But farmed fish might not ever be a sustained replacement for caught



even higher, at 3.7 million tonnes of meal and 840,000 tonnes of oil. Although the proportion of wild caught fish fed to farmed fish is small when set against the global total of capture fisheries, it is still a sticking point if the rationale for fish farming is to relieve pressure on overfished oceans.

### Strange chickens

Of the top seven most heavily farmed fish by weight, five are carp species, which generally require less food supplementation than other species. But carp is generally unfamiliar to Western diets. Number eight on the list has been quickly rising up the ranks and in the consciousness of Westerners: *Oreochromis niloticus*, the Nile tilapia. Tilapia — affectionately dubbed the aquatic chicken for their speedy and efficient growth — are to many a nearly perfect aquaculture species. The fish grow fast and are not choosy about where they live or what they eat. They occupy a low position on the food chain, so there's little opportunity for mercury to build up in their flesh — as is the case for some predatory species — and their flavour is sweet and inoffensive.

Although high-capacity tilapia farms once led to rapid breeding and disappointingly small adults,

interbreeding species or using hormones can create single-sex broods, effectively taking the brakes off the population's rapid growth. Production boomed from next to nothing in the 1970s and 1980s to more than 2 million tonnes by 2007.

But much like the land-based avian namesake, these aquatic chickens are sneered at by many fish fans for their bland ubiquity. Westerners favour animals higher up the food chain: salmon, tuna and striped bass, among others. These animals have more rarefied tastes than the grubbing tilapia, and as they fetch a high price, farmers are keen to keep them on the menu.

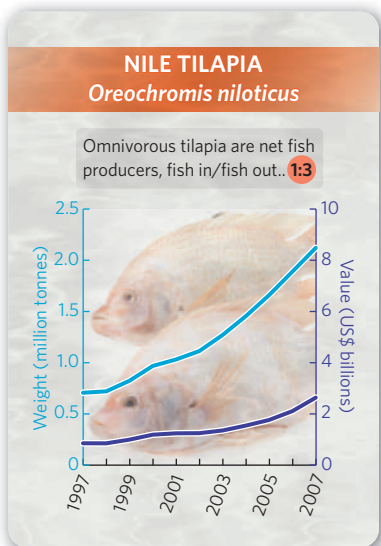
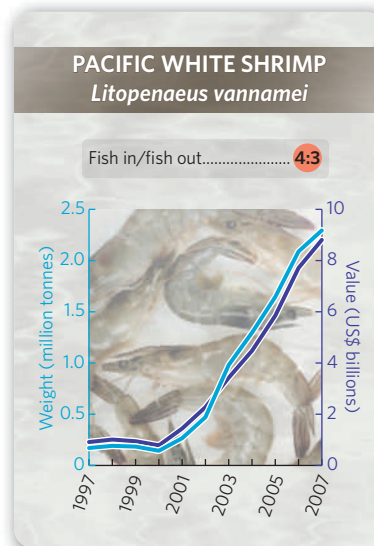
To this end, some of the sea's carnivores have adopted a more vegetarian-like diet. In many farms, salmon, the predator farmed most heavily, have become accustomed to a diet of at least

25% soya bean, supplemented with fishmeal and fish oil at crucial times in their growth. "Even though they look like fish they are almost more similar to pigs," says Carlos Duarte of the Mediterranean Institute for Advanced Studies in Mallorca, Spain.

Other salmon may eat even more soya beans but such diets will reduce the levels of healthy omega-3 oils that have been a major selling point for salmon. Solutions in the pipeline include soya-bean crops that have been genetically modified to put these oils back in, and possibly even modifying the

fish themselves. Traits such as fast maturation can be introduced through selective breeding. Other traits, such as disease resistance, will be harder to obtain without resorting to genetic modification, says Eric Hallerman, head of the Department of Fisheries and Wildlife Sciences at Virginia Tech in Blacksburg.

The US Food and Drug Administration (FDA) is considering approval of an Atlantic salmon that has been modified with the gene encoding for growth hormone from a Pacific Chinook salmon (*Oncorhynchus tshawytscha*). The addition of the gene could cut the time taken to reach market size by between a third and a half. The decision has been in the works for a long time<sup>4</sup>, however, and there are no indications of a quick resolution. "I'd be really surprised if the FDA went forward before we give in our final report [on ecological risk assessment], which will be next September or October," says Hallerman, who is leading the work.



M. SCHUTT/DPA/CORBIS

P. DALE/ALAMY





D. CRESSEY

And there are additional costs to consider. The big issue here, as with all genetically modified organisms, is preventing the fish escaping into the wild, where they may outcompete local fish, or where the modified genes could transfer to wild stocks.

To avoid this, Hallerman says that the modified salmon should only be grown onshore in tanks. At present, nearly all farmed salmon are grown in cages at sea, which puts onshore farmers at a disadvantage.

However, some in the fish-farming community say that this is the best place to farm fish. By using a series of graduated tanks — small tanks for hatchery up to large 'grow out' tanks — and recirculating water with pumps it is possible to grow dinner inside what is basically a warehouse. It also removes the variability that comes with being open to the ocean, allowing every stage of a fish's development to be carefully monitored and controlled, and helping to ensure a healthy fish for the market.

Although such technology costs money, some people using these systems are already making money, and any toughening up of regulations for outdoor aquaculture could make the recirculation systems even more attractive.

"Twenty years ago people would have told you we can't raise fish on land in recirculation systems. Ten years ago they would have said we can't raise them and make a dollar," says Hallerman. "People laughed at me 10 years ago when I said this, they didn't take it seriously. Now they're like 'show me how it works on a spreadsheet.'"

But many people think the real opportunities lie in the deep ocean. Most fish farms are confined to the narrow strip of water near the coast,

but advocates of 'open ocean' farming say that going farther out to sea is the way forwards. "On the one side there is unlimited potential," says James Diana, from the School of Natural Resources and Environment at the University of Michigan in Ann Arbor, "but it's also very expensive. It is a question of which of those two will win out."

### Out to sea

Going offshore would remove many of the problems of near-shore farms: water quality is generally higher and there are fewer conflicts with recreational water users. But the open ocean can be fierce, and farms will need to be engineered much more heavily than they are now. Then there is the tricky problem of licensing. In the United States, for instance, there is no regulatory system for licensing fish farms in federal waters, so farms are limited to the 3 nautical miles (around 5.6 kilometres) off-shore that fall under state control.

"We currently do not have a regulatory framework for issuing permits for aquaculture in federal waters," says Rubino. "There are a few companies using offshore technologies in open ocean conditions in state waters, but none in offshore (federal) waters," he explains. "It's not suitable for all species and all locations but it's certainly something that many have recommended we look at going forward."

Although proponents think that being out at sea may cause less problems than being close to shore, some environmentalists have been

fiercely opposed to open-ocean systems<sup>5</sup>.

For Duarte, however, the opponents are missing the greater picture of food production. "We don't need to occupy a major fraction of the oceans to grow sufficient food. We have transformed 50% of the surface of the continents into crop lands and graze lands and yet we can probably [make] do with much less than 10% of the surface of the coastal ocean."

Consumer demand will probably push the technology as far as it can go, and in the West that means focusing on marine predators.

But Duarte says that could change with the global food shortages. "Then the demand will be for mass production of food, and not so much for specific elements of quality."

The farm at Howietoun will still probably be producing trout in 20 years. The restaurant nearby may still be

serving farmed fish, be it genetically modified salmon, humble tilapia, tank-farmed tuna, or even carp.

If those who are partial to the taste of fish are lucky, an extra 30 million tonnes of fish will be on the market. Dinner won't be coming from the sea though, at least not in the traditional sense. The future is farmed.

**Daniel Cressey writes for *Nature* from London.**

**"Consumer demand will be for mass production of food, and not so much for specific elements of quality."**

— Carlos Duarte

1. FAO *The State of World Fisheries and Aquaculture 2008* (FAO, 2009).
2. FAO *Opportunities for Addressing the Challenges in Meeting the Rising Global Demand for Food Fish from Aquaculture* (FAO, 2008).
3. Tacon, A. & Metian, M. *Aquaculture* **285**, 146–158 (2008).
4. *Nature* **406**, 10–12 (2000).
5. *Nature* **431**, 502–504 (2004).



# THE ROCK THAT FELL TO EARTH

When an asteroid was spotted heading towards our planet last October, researchers rushed to document a cosmic impact from start to finish for the first time. **Roberta Kwok** tells the tale.

**A**round midnight on 6 October 2008, a white dot flitted across the screen of Richard Kowalski's computer at an observatory atop Mount Lemmon in Arizona. Kowalski had seen hundreds of such dots during three and a half years of scanning telescope images for asteroids that might hit Earth or come close. He followed the object through the night and submitted the coordinates, as usual, to the Minor Planet Center in Cambridge, Massachusetts, which keeps track of asteroids and other small bodies. When the sky began to brighten, he shut down the telescope, went to the dorm down the mountain and fell asleep.

The only thing that had puzzled Kowalski about the midnight blip was the Minor Planet Center's response to his report. Its website posted the discovery right away but when he tried to add more data, the system stayed silent.

Tim Spahr, the Minor Planet Center's director, found out why the following morning. The centre's software computes orbits automatically, but this asteroid was unusually close to Earth. "The computer ran to me for help," says Spahr. He did some quick calculations on Kowalski's data to figure out the path of the asteroid, which was now named 2008 TC<sub>3</sub>. "As soon as I looked at it and did an orbit manually, it was clear it was going to hit Earth," he says.

The brightness of 2008 TC<sub>3</sub> suggested it was only a few metres across and, assuming it was a common rocky asteroid, would probably split into fragments soon after entering the atmosphere. But safe as that might seem, Spahr had procedures to follow. He called Lindley Johnson, head of NASA's Near Earth Object

Observations programme in Washington DC, on his BlackBerry — a number only to be used in emergencies.

"Hey Lindley, it's Tim," said Spahr. "Why would I be calling you?"

Johnson's response: "We're going to get hit?"

Spahr also called astronomer Steve Chesley of the Jet Propulsion Laboratory (JPL) in Pasadena, California, who at the time was hustling his kids out of the door for school. Chesley hurried into the office, ran a program to calculate the asteroid's orbit and "was astounded to see 100% impact probability," he says. "I'd never seen that before in my life." Chesley calculated that the asteroid would hit Earth's atmosphere

less than 13 hours later, at 2:46 UT the next day; the impact site would be northern Sudan, where the local time would be 5:46 a.m.. He sent his results to NASA headquarters and the Minor Planet Center, which circulated an electronic bulletin to a worldwide network of astronomers. A group called NEODys in Pisa, Italy, also confirmed that an impact was nearly certain.

Although several small objects such as 2008 TC<sub>3</sub> hit Earth each year, researchers had never spotted one before it struck. Kowalski's discovery, therefore, provided a unique chance to study an asteroid and its demise in real time, if astronomers could mobilize resources around the world quickly enough.

Soon e-mails and phone calls were flying across the globe as scientists raced to coordinate observations of the incoming asteroid. "IMPACT TONIGHT!!!" wrote physicist Mark

Boslough of Sandia National Laboratories in Albuquerque, New Mexico, to colleagues, including a Sandia engineer responsible for monitoring US government satellite data.

## Countdown to impact

Peter Brown, an astronomer at the University of Western Ontario in Canada who heard the news from JPL, ran to his local observatory, fired up the telescope and began tracking the asteroid, which looked like "a very small, faint, fast-moving streak," he says. Alan Fitzsimmons at Queen's University Belfast in Northern Ireland called two of his colleagues, who had just arrived at the William Herschel Telescope at La

Palma on the Canary Islands and were not scheduled to use the telescope until the next day.

"Listen guys, this is happening, this is going to happen tonight," he told the researchers, who arranged to borrow an hour of observing time from

**"Listen guys, this is happening, this is going to happen tonight."  
— Alan Fitzsimmons**

another astronomer.

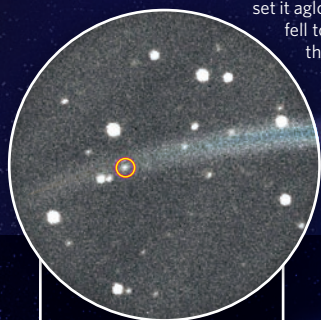
All day, observations poured into the Minor Planet Center, which released new data and orbit calculations several times an hour. NASA notified other government agencies, including the state and defence departments, and issued a press release that afternoon saying that the collision could set off "a potentially brilliant natural fireworks display". About an hour before impact, the asteroid slipped into Earth's shadow and out of view to optical telescopes. By then, astronomers from 26 observatories worldwide had already captured and submitted about 570 observations, allowing JPL to refine





# A 2008 TC<sub>3</sub> SPACE ODYSSEY

The little boulder 2008 TC<sub>3</sub> went through a series of name changes during its brief moment in the scientific spotlight. In space, the hunk of rock was called an asteroid or meteoroid. After it hit Earth's atmosphere, frictional heating set it aglow and it became a meteor. The pieces that fell to the ground are called meteorites. Here is the 2008 TC<sub>3</sub> biography, from the moment it was discovered.

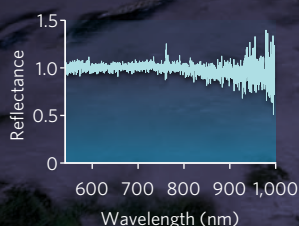


6 OCT 2008  
06:39 UT

A fast-moving meteoroid close to Earth was spotted by the Catalina Sky Survey on Mount Lemmon in Arizona. Orbital calculations suggested it would hit the planet in 20 hours.

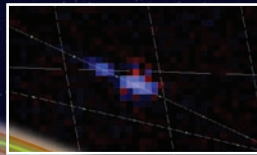
6 OCT 2008  
22:22–22:28 UT

When the meteoroid was 121,100 kilometres from Earth, a telescope in the Canary Islands measured how much light the body reflected at different wavelengths.



7 OCT 2008  
02:45:46 UT

When the meteoroid broke apart, it left behind clouds of hot dust, observed by the Meteosat-8 weather satellite.



7 OCT 2008  
03:27 UT

A photograph captured clouds left behind after the fireball disappeared.



7 OCT 2008  
02:45:40 UT

Ron de Poorter, a KLM pilot flying at an altitude of 10,700 metres over Chad, saw three or four short pulses of light beyond the horizon as the meteoroid flared through the sky.



DECEMBER  
TO MARCH

A search team combed the desert multiple times and recovered some 280 meteorites.



NASA/UNIV. ARIZONA; FULL MOON PHOTOGRAPHY; A. FITZSIMMONS; QUEEN'S UNIV. BELFAST

its predicted collision time to 2:45:28 UT, give or take 15 seconds.

As the countdown progressed, Jacob Kuiper fretted. Kuiper, an aviation meteorologist on the night shift at the Royal Netherlands Meteorological Institute in De Bilt, had seen an e-mail about the incoming asteroid. And he was worried that no one would see the explosion in the sparsely populated Nubian Desert.

With less than 45 minutes left, Kuiper realized he could notify Air France-KLM — the airline to which he routinely issued weather reports — which probably had planes flying over Africa. About ten minutes later, pilot Ron de Poorter received a message print-out in the cockpit of KLM flight 592, flying north from Johannesburg to Amsterdam. The message gave the latitude and longitude of the predicted asteroid impact. De Poorter calculated that he would be a distant 1,400 kilometres from the collision. Still, at the appointed time he and his co-pilot dimmed the instrument lights and peered northeast.

Far above the plane, asteroid 2008 TC<sub>3</sub> hit the top of the atmosphere at about 12,400 metres per second. The collision heated and vaporized the outside of the rock, ripping material from its surface. The impact of rock atoms with air molecules created a brilliant flash that lit the desert below. Less than 20 seconds after 2008 TC<sub>3</sub> entered the atmosphere, calculations suggest, pressure on the rock triggered a series of explosions that shattered it, leaving a trail of hot dust.

From the cockpit of his plane, de Poorter saw

flickerings of yellowish-red light beyond the horizon, like distant gunfire. The flash woke a station manager at a railway outpost in Sudan. In a village near the Egyptian border, people returning from morning prayers saw a fireball that brightened and flared out, according to accounts collected later by researchers.

Electronic eyes watched, too. US government satellites spotted the rock when it was 65 kilometres above the ground. Moments later, it was picked up by a European weather satellite, which caught two dust clouds and light from the fireball. An array of microbarometers in Kenya normally used to monitor for nuclear explosions detected low-frequency sound

waves from the blast, which Brown later calculated would be equivalent to about 1–2 kilotonnes of TNT, roughly one-tenth the size of the atomic bomb dropped on Hiroshima.

Tracking of the fireball's trajectory by US satellites showed that JPL accurately predicted the object's location within a few kilometres and a few seconds. "We have never had such a concrete affirmation that all the machinery works," says Chesley.

But for Peter Jenniskens, an astronomer at the SETI Institute in Mountain View, California, the spectacular light show was not enough. For weeks after the asteroid hit, Jenniskens, who studies meteor showers, waited to hear whether someone had found the fallen meteorites. No news emerged. "Somebody needed to do something," he says.

Jenniskens flew to Sudan in early December and met with Muawia Hamid Shaddad,

an astronomer at the University of Khartoum who had already obtained pictures of the fireball's trail from locals. Together, they drove north from Khartoum to the border town of Wadi Halfa, asking villagers where the fireball had exploded in the sky. These eyewitness accounts convinced Jenniskens that the rock had disintegrated high in the atmosphere — in good agreement with US satellite data — and that any fragments were most likely to be found southwest of Station 6, a tiny railroad outpost in the Nubian Desert.

## Desert search

On 6 December 2008, Jenniskens and Shaddad set out with a group of 45 students and staff from the University of Khartoum to scour the area. Team members lined up about 20 metres apart over a kilometre-wide strip, facing a sea of sand and gravel interspersed with hills, rocky outcrops and dry winding riverbeds. Flanked by two pairs of cars and trailed by a camera crew from news network Al Jazeera, the line of searchers began marching slowly east, like the teeth of a massive comb being dragged through the desert.

Towards the end of the day, a car approached Jenniskens with news that a student might have found a meteorite. "I remember thinking, 'oh no, not again,'" says Jenniskens, who had already fielded several false alarms. Still, he jumped in the car and drove to the student, who presented him with a small square fragment, about a centimetre and a half across with a thin, glassy outer layer. The surface resembled the crust that meteorites form after being melted and solidified, and the rock's deep black colour suggested it was freshly

**"We have never had such concrete affirmation all the machinery works."**  
— Steve Chesley



fallen. It was the team's first meteorite — and the first time that scientists had ever recovered a meteorite from an asteroid detected in space (see page 485).

The next day, the team walked 8 kilometres and found 5 meteorites, all very dark and rounded. On the third day, a trek of 18 kilometres yielded larger meteorites nearly 10 centimetres across. A few weeks later, a team of 72 students and staff found 32 more, and the most recent field campaign, completed in March, brought the tally to about 280 fragments weighing a total of several kilograms.

Jenniskens couriered a sample to Mike Zolensky, a cosmic mineralogist at the NASA Johnson Space Center in Houston, Texas. Examining the rock, Zolensky discovered that it contained large chunks of carbon and glassy mineral grains resembling sugar crystals. Tests at other labs confirmed that the sample was a ureilite, a type of meteorite thought to come from asteroids that have melted during their time in space. Only 0.5% of objects that hit Earth yield fragments in this category. But 2008 TC<sub>3</sub>'s pieces are strange even for ureilites: they are riddled with an unusually large number of holes, says Zolensky. "It boggles the mind that something that porous could survive as a solid object," he says.

The findings suggest that 2008 TC<sub>3</sub> broke from the surface of a larger asteroid, as the pores would have been crushed if they were near the rock's centre, says Zolensky. He suggests that future studies of the meteorites' chemistry could help reveal the history of its parent asteroid. Moreover, the new finds might eventually yield clues to how planets form, he says, because the asteroid had melted during its history, a process that young planets go through.

2008 TC<sub>3</sub> gave astronomers a rare chance to connect a dot in the sky with rocks in their hands. "We have a lot of meteorites on the ground and a whole lot of asteroids up there, and forging a link is not easy," says Don Yeomans, manager of NASA's Near-Earth Object Program Office at JPL.

Jenniskens and his team concluded the asteroid belonged to a group called F-class asteroids. These asteroids reflect very little light, and scientists had been unsure what they were made of. The new evidence "opens a huge window", says Glenn MacPherson, a meteorite curator

at the Smithsonian Institution in Washington DC, who was not involved in the studies of 2008 TC<sub>3</sub>. Although not all F-class asteroids may be the same, he says, the data suggest at least some of them may contain the same material as ureilites, such as carbon and iron.

Clark Chapman, a planetary scientist at the Southwest Research Institute in Boulder, Colorado, says the connection between F-class asteroids and ureilites does not surprise him. But, he adds, "this is a proven link and we don't have many of those".

Scientists have tried to figure out the composition of asteroids by studying how they reflect various wavelengths of light and matching these features to meteorite samples in the lab. But such connections are often tenuous unless the reflection signature is very distinct. The most secure example is an asteroid called 4 Vesta, which has been associated with a group of igneous meteorites. No missions have yet returned asteroid fragments to Earth, although a NASA spacecraft orbited the asteroid Eros for a year and landed on it in 2001. Japan's Hayabusa mission attempted to collect a sample from the asteroid Itokawa in 2005; scientists will find out whether it succeeded when the spacecraft returns next year.

Knowing what asteroids are made of will be crucial if we ever need to deflect one, says Yeomans.

NASA aims to provide decades of warning if any killer asteroids are headed for Earth so that a strategy can be devised

to avoid a collision. That strategy will differ for various asteroids, which can range from "wimpy ex-cometary fluffballs", to solid rock, to slabs of nickel-iron, says Yeomans.

With the advent of new surveys, scientists could spot objects hurtling towards Earth more frequently. Today's surveys have found almost 90% of near-Earth objects with a diameter of 1 kilometre or larger, says Yeomans, but smaller rocks can easily slip by unnoticed. Discovering 2008 TC<sub>3</sub> was like finding

"a man in a dark grey suit 50% farther away than the Moon", says Kowalski, who is part of the Catalina Sky Survey, an effort that discovers 70% of all the near-Earth objects found every year. The detection rate will increase with the next generation of surveys, per-

haps up to a few Earth-bound asteroids per year, says Alan Harris, a planetary astronomer at the Space Science Institute who is based in La Canada, California. The Panoramic Survey Telescope and Rapid Response System (Pan-STARRS) in Hawaii will officially begin observations with its prototype system this year, and the Large Synoptic Survey Telescope in Chile is scheduled to begin full operations in 2016.

In the meantime, Kowalski and his colleagues are still on the job. The night after spotting asteroid 2008 TC<sub>3</sub>, Kowalski headed back up Mount Lemmon, heated his dinner and settled down in the telescope's control room. As his discovery plunged towards the desert on the other side of the world, Kowalski was surveying another part of the sky, waiting for the next white dot.

**Roberta Kwok is a news intern in Nature's Washington DC office.**

**"It's like finding a man in a dark grey suit 50% farther away than the Moon"**  
— Richard Kowalski



Peter Jenniskens (above) led the search for meteorite fragments in the Sudan desert (inset).

P. JENNISKENS

# CORRESPONDENCE

## We need to tackle the mismatch between supply and demand

SIR — Your Editorial 'Crisis of confidence' (*Nature* **457**, 635; 2009), about US graduates pursuing careers in biomedical research, hits the nail on the head. The problem is not that junior faculty careers are not fostered. Their crisis is predominantly due to perennial mismatches between supply and demand. The aura of an unstable career permeates the entire profession and has diverted the best students from seeking careers in biomedical sciences.

Strategically, there are two main issues facing the funding of US biomedical research, especially through the National Institutes of Health (NIH).

First, we all wish to see continued growth at this time of exciting developments in biomedical research. This goal seems to enjoy wide public support, and the new stimulus budget reflects a deep commitment to funding science. We need to convince all who will listen of the importance of this goal.

Second, we should be able to work with the hand that is dealt us: annual budgets from Congress. Why can't the NIH evolve mechanisms to regulate the impact of the upswings (increased congressional allocation) and downturns (periods of stagnant growth)? This is a key goal, as ensuring reliable growth is the best mechanism for getting new blood into biomedical research.

Here are some suggestions to consider. One is to modulate the amount of funding and grants depending on funding. Supplements can be used during periods of growth and taken away during downturns, which will tend to directly stabilize the number of grants, and thus the number of applicants and their success rates.

Another suggestion is to adjust the effort a principal investigator can charge to grants. This can be decreased during expansion; universities have to contribute

more. This would impose on departments and universities a clear restraint on uncritical runaway hiring. Also, during periods of growth, funding should be diverted to infrastructure and programme activities: instrumentation, renovations and collaborative grants.

Approaches such as these and others, in aggregate, would reduce the incentive for rapid growth during periods of budgetary expansion at the NIH and would redirect funding to the R01 (research project grant) pool during periods of stagnation. Hence, they are likely to result in more 'steady state' growth that will reinvigorate career development in the biomedical and life sciences.

**Yusuf A. Hannun** Department of Biochemistry and Molecular Biology, Medical University of South Carolina, Charleston, South Carolina 29425, USA  
e-mail: hannun@musc.edu

## Review: important to prevent a return to abuses of the past

SIR — As a former chair of an institutional review board (IRB), I sympathize with the concern expressed in Scott Kim and colleagues' Commentary 'Pruning the regulatory tree' (*Nature* **457**, 534–535; 2009) about the heavy burden imposed by dealing with exempt research. Unfortunately, their fix won't work.

Many years ago, before the US National Institutes of Health clamped down, exemption was handled very casually at many institutions, even to the extent of researchers declaring themselves or their students exempt without any formal review. The result was a widespread culture of non-compliance with the common rule, with many dubious and improper claims of exemption. That is why IRBs at some institutions do not grant them — exemptions got a bad name.

A new regulation that "exempts minimal-risk research from IRB

review" would certainly send a clear and unambiguous message: an invitation to return to the abusive practices of the past. If an IRB chair or a full review board feels the need to clamp down on exemption declarations to help researchers remain sensitive to their obligations to protect their human subjects, please give them your understanding and support. They are doing the right thing.

Kim's comments in *Nature Network* (<http://tinyurl.com/am7l24>) include an example of financial regulations that may seem too fussy. He is right on target with his choice of domain from which to draw analogies. I think most would agree that the streamlining of regulations in the financial industry — the undoing of rules requiring transparency and accountability, record-keeping and review — tempted people into excesses that have now severely damaged our economy. Let us not make a similar mistake with IRBs.

Without truly independent review of research protocols involving human subjects, at best preventable mistakes will happen because nobody looked and, at worst, significant harm may result.

**Herbert J. Bernstein** Department of Mathematics and Computer Science, Dowling College, Kramer Science Center, Oakdale, New York 11769, USA  
e-mail: yaya@dowling.edu

## Review: necessary for protection even in minimal-risk research

SIR — Although I am sympathetic to the suggestion made in Scott Kim and colleagues' Commentary (*Nature* **457**, 534–535; 2009) that we deregulate minimal-risk research by treating it as exempt from assessment by institutional review boards (IRBs), I do not believe that their proposal provides adequate protection for human subjects.

Most institutions require their researchers to submit a short application to the IRB requesting

an exemption, but once an exemption is granted, the IRB need not ever see the research again (Center for Advanced Study *Improving the System for Protecting Human Subjects: Counteracting IRB 'Mission Creep'*; CAS, 2006). The problem with treating minimal-risk research as exempt is that such studies are not risk-free. Researchers and institutions need to ensure that risks are adequately addressed. An independent body such as the IRB can make useful suggestions for minimizing risks, protecting confidentiality and interacting with human subjects. It can also provide continuing review and supervision of research, which is valuable when unanticipated problems arise or a study changes course.

Although minimal-risk studies do not call for as much scrutiny as more risky ones, they still require some independent overview. Minimal-risk social-science research, for example, can involve questioning people about sensitive topics such as sexuality, domestic violence or illicit drug use. And in minimal-risk biomedical research, risks may be associated with protecting the confidentiality of genomic samples or data.

A wiser proposal, which has received some attention in the literature, would be to make better use of the expedited review option for minimal-risk research; indeed, the CAS study above notes that IRBs are not making adequate use of this option. In expedited review, the IRB chair or another designated person conducts the review. Studies that are approved are subjected to the same review criteria that are applied to studies reviewed by the full board, as well as receiving continuing review. Expedited review can reduce the administrative burden without compromising the rights or welfare of human subjects.

**David B. Resnik** National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, North Carolina 27709, USA  
e-mail: resnikd@niehs.nih.gov



## COMMENTARY

# The dangers of diagnostic monopolies

In the first of two commentaries on intellectual property, **Robert Cook-Deegan**, **Subhashini Chandrasekharan** and **Misha Angrist** show how the United States can address glitches with exclusive licences.

Gene patents are meant to encourage innovation, but in DNA diagnostics, they have stirred controversy. Amid worries and worst-case scenarios, there are few empirical studies to help form an accurate picture of how patents affect clinical genetic testing in the United States. We assembled eight case studies addressing the effects of patents and licensing on access to genetic tests for ten conditions (see 'Gene-test licensing in the United States', overleaf). The studies were prepared for a task force of the Secretary's Advisory Committee on Genetics, Health and Society and are available as inputs to a draft report<sup>1</sup> for the US Secretary of Health and Human Services. They paint a complex picture of the patent landscape.

Despite the fears, patents have not caused irreparable harm in genetic diagnostics, but neither have they proven greatly advantageous. Although our findings detect no pervasive effects that consistently help or hinder clinical access to genetic testing, there are some problems that could be addressed to the benefit of patients, researchers, health professionals and companies alike.

## Pricing and availability

Most concerns centre on monopoly situations, in which exclusive licensing results in a single dominant provider. But prices of patented and exclusively licensed tests are not dramatically or consistently higher than those of tests without a monopoly — a contrast with the strong price effects of drug patents. For example, unit prices for BRCA testing (for breast cancer susceptibility) — provided solely in the United States by Myriad Genetics in Salt Lake City, Utah — are comparable with similar tests for colon cancer susceptibility available from many labs under non-exclusive licences. We also do not find consistent price effects of patents in other case studies. Other factors affect pricing, such as efficiencies of scale from high-volume testing and the way health plans use administrative codes to pay for tests.

Prices would matter less if everyone in the United States were insured for genetic diagnostic tests. Health-plan coverage and reimbursement are problems even when many labs offer a test. But with only one provider,

the absence of alternatives can exacerbate the problem. Some health plans do not have a contract with the sole provider, for example. Nevertheless, after ten years of testing, Myriad now reports payment arrangements that substantially cover costs for more than 90% of tests, so problems in coverage and payment cannot solely be attributable to monopolies.

One justification for gene patents is that they speed up the development of tests. But the patent incentive is usually not necessary. Barriers to test development are fairly low compared with in the 1990s, when the genetic bases for most of the conditions we studied were elucidated. Academic labs typically offer testing soon after publication of an association, when demand exists. After exclusive licences are issued, however, the licensee enforces patents to 'clear' the market of competitors.

Monopoly effects on test quality are equivocal. For example, in 2006, Myriad's

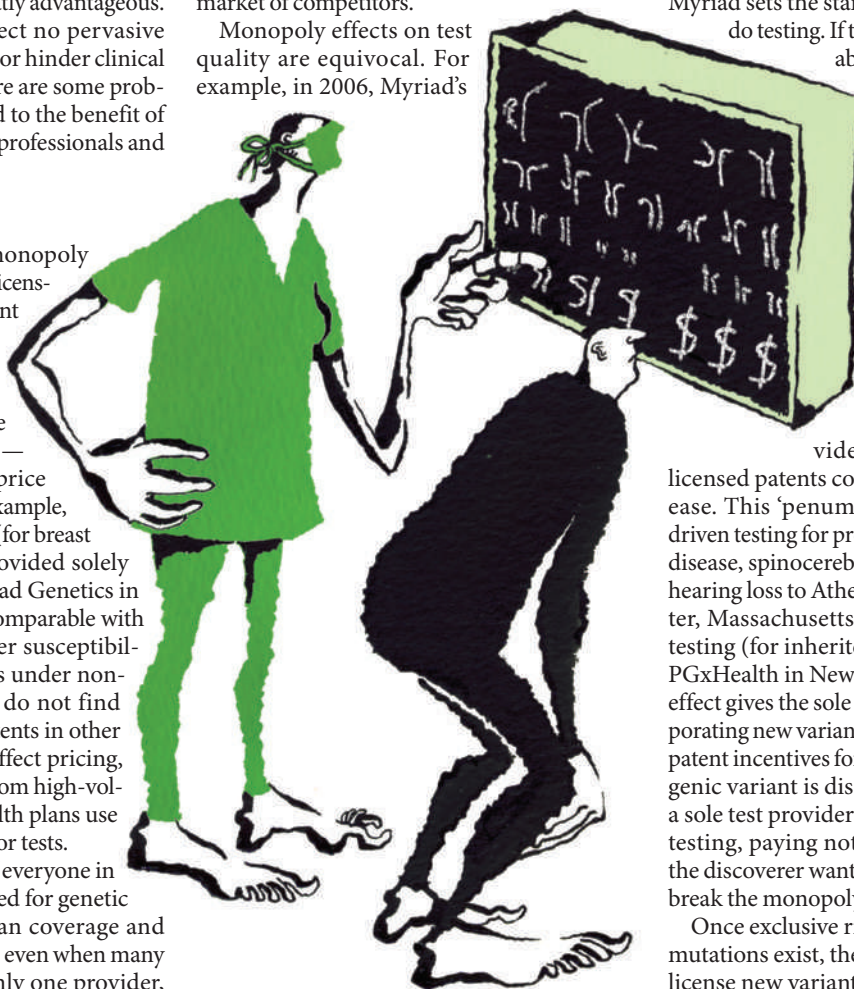
methods of BRCA testing were shown to miss some DNA deletions and rearrangements<sup>2</sup>. Yet such problems cannot be ascribed only to the monopoly. Problems are apparent in genetic testing for other conditions offered by labs with non-exclusive rights. Test quality is a general problem but monopolies can exacerbate it.

Sole providers can also, for better or worse, establish standards of care. Licences for patents related to risk prediction for Huntington's disease and Alzheimer's disease enforced clinical standards set by the professional community<sup>3</sup>. If compliance with standards is desirable, patents can achieve it. But is a standard of care set by a single provider desirable? In other countries, testing labs take various approaches before ordering full-sequence BRCA testing as licensed by Myriad. Yet in the United States, Myriad sets the standard because it alone can do testing. If testing alternatives are valuable, exclusive licences limit their availability.

Exclusive licences are difficult for other companies to develop around. Most genetic disorders are heterogeneous — different genes and mutations lead to clinically similar syndromes. An exclusive licence to one or a few common genes or mutations can drive testing to one provider, regardless of whether

licensed patents cover all varieties of the disease. This 'penumbra effect' seems to have driven testing for predisposition to Alzheimer's disease, spinocerebellar ataxia and hereditary hearing loss to Athena Diagnostics in Worcester, Massachusetts, and long-QT syndrome testing (for inherited cardiac arrhythmia) to PGxHealth in New Haven, Connecticut. The effect gives the sole provider a leg-up in incorporating new variants into its assay. It also alters patent incentives for others. If a gene or pathogenic variant is discovered but not patented, a sole test provider can incorporate it into its testing, paying nothing to the discoverer. If the discoverer wants a piece of the action or to break the monopoly, he or she must patent it.

Once exclusive rights to common genes or mutations exist, the only realistic option is to license new variants exclusively, either to the



B. MELLOR



GENE-TEST LICENSING IN THE UNITED STATES			
Condition	Test provider(s)	Patent(s) owner	Licensing
Tay-Sachs disease	Various	National Institutes of Health (NIH)	Not licensed
Inherited risk of colorectal cancer	Myriad and others	University	Non-exclusive
Inherited risk of breast/ovarian cancer	Myriad Genetics	University, Myriad, NIH	Myriad (exclusive)
Canavan disease	Various	Miami Children's Hospital	Private settlement
Cystic fibrosis	Various	University and hospital	Non-exclusive
Alzheimer's disease	Athena Diagnostics	University	Athena (exclusive)
Spinocerebellar ataxia	Athena	University and Athena	Athena (exclusive)
Haemochromatosis	Various	BioRad	Non-exclusive
Hearing loss	Athena and others	University and hospital	Athena (exclusive)
Long-QT syndrome	PGxHealth	University	Possible mutual block

existing sole provider (augmenting the monopoly) or to a rival company. Indeed, the potential for mutually blocking patent rights seems to be developing: PGxHealth has been the sole long-QT test laboratory for several years, based on exclusive rights to several patents. But the University of Utah in Salt Lake City recently awarded exclusive rights for other long-QT-related patents to Bio-Reference Laboratories in Elmwood Park, New Jersey. It remains to be seen how this situation will affect patient access.

### An instrumental right

Some companies that have used patents to build their businesses claim that their services are of equal or better quality than university and reference laboratory services. But, in our case studies, clinicians, patient groups and even other companies argue that patents should be non-exclusively licensed for diagnostic purposes. Who should provide proof as to whether patents and exclusive licences are promoting or interfering with progress?

We propose that patient rights should trump patent rights if a company engages in practices that undermine the purpose of patents in the first place (see 'Recommendations in cases of monopoly'). The burden of proof should lie with the companies enforcing patent rights.

Patenting is a right, but an instrumental one. The US Constitution states that patents are granted to promote public good through advancing science and technology. The stakes are high when human health is involved. Governments should take action when harm is apparent or foreseeable, especially when technologies spring from public funding. The government has powers that have not been used: to decide coverage and reimbursement for tests, to regulate and to ensure that patent rights promote health and safety under the 1980 Bayh-Dole Act.

Academic institutions play an important part in clinical genetic testing. They own most of the

patents relevant to Mendelian disease testing<sup>4</sup>, and 60% of clinical genetic testing laboratories are within universities<sup>5</sup>. Academic institutions thus both own most genetic-diagnostic patents and operate many of the laboratories against which such patents are enforced. This paradox derives from technology licensing and clinical laboratory services that are run by different parts of universities and have different missions. These need to be aligned. A non-binding statement endorsed by the Association of University Technology Managers in 2007 says "licenses should not hinder clinical research, professional education and training, use by public health authorities, independent validation of test results or quality verification and/or control"<sup>6</sup>.

As we attempted to reconstruct custody chains for the relevant intellectual property, university technology licensing offices were often among the missing links, failing to provide information despite repeated inquiries. If we are to demand transparency from private genetic testing laboratories, academic institutions that license technologies arising from federal grants should be at least as willing to

provide information. Research institutions that use public dollars to create licensed inventions are publicly accountable for the disposition of the resulting intellectual property.

Looking to the future, robust genomic technologies promise to transform genetic testing. The price of full-genome sequencing will drop, and speed and accuracy will improve. Soon, sequencing a person's entire genome will cost less than current tests for one or a few genes. Yet thousands of patents claim human DNA sequences<sup>7</sup> and so some patent claims will be infringed by full-genome sequencing. This legacy of patented sequences could cause considerable mischief, depending on how intellectual property is managed. As genetic testing is transformed by new technologies, patenting and licensing practices and government oversight should focus on the net social benefit for patients as well as on freedom to innovate. ■

**Robert Cook-Deegan, Subhashini Chandrasekharan and Misha Angrist**, are at the Institute for Genome Sciences and Policy, Duke University, Durham, North Carolina 27708, USA. e-mail: gelp@duke.edu

1. Secretary's Advisory Committee on Genetics, Health, and Society *Public Consultation Draft Report on Gene Patents and Licensing Practices and their Impact on Patient Access to Genetic Tests* (2009). Available at [http://oba.od.nih.gov/SACGHS/sacghs\\_public\\_comments.html](http://oba.od.nih.gov/SACGHS/sacghs_public_comments.html)
2. Walsh, T. et al. *J. Am. Med. Assoc.* **295**, 1379-1388 (2006).
3. *Reaping the Benefits of Genomic and Proteomic Research: Intellectual Property Rights, Innovation, and Public Health* (Nat'l Acad. Press, 2005).
4. Henry, M. R., Cho, M. K., Weaver, M. A. & Merz, J. F. *J. Law. Med. Ethics.* **31**, 442-449 (2003).
5. Cho, M. K. et al. *J. Mol. Diagn.* **5**, 3-8 (2003).
6. *In the Public Interest: Nine Points to Consider in Licensing University Technology* (Assoc. Univ. Technol. Managers, 2007). Available at [www-leland.stanford.edu/group/OTL/industry/resources/whitepaper-10.pdf](http://www-leland.stanford.edu/group/OTL/industry/resources/whitepaper-10.pdf)
7. Jensen K. & Murray F. *Science* **310**, 239-240 (2005).

This article represents the authors' opinions and not those of the Secretary's Advisory Committee for Genetics, Health and Society or its task force.

See Editorial, page 386, and online at <http://tinyurl.com/dlnksh>.

### Recommendations in cases of monopoly

We find six obstructive practices and policies that we think should negate gene patent enforcement for diagnostic-testing companies with exclusive rights. Patient rights should trump patent rights if a company's policies do not permit:

- Basic and clinical research, including genetic testing in clinical trials or health-services research.
- Performing a test in a form that it does not offer (such as prenatal or preimplantation diagnosis).
- Testing in a territory where the company does not offer a test but has exclusive rights.
- Getting second opinions or verification testing.
- Testing those not covered by its payment agreements with insurers and health plans.
- Research and development to make testing more comprehensive, more accurate or less expensive.

As unique guardians of key information, sole-provider laboratories have a responsibility to:

- Share data about allele frequency, number of tests, aggregate results and other facts relevant to public health.
- Contribute to public databases that catalogue variants and contribute to their interpretation.
- Perform proficiency testing and otherwise ensure quality of testing.
- Disclose intellectual-property ownership and licensing provisions related to genetic diagnostics.

## COMMENTARY

# The phantom menace of gene patents

In this, the second of two Commentaries, **Sibylle Gaisser, Michael M. Hopkins** and colleagues discuss a survey demonstrating that European health-care systems are ill prepared for the commercial reality of gene patents.

In 1998, the European Parliament passed a law that requires EU Member States to recognize isolated genes and nucleotide sequences as patentable inventions, further reiterating obligations under the European Patent Convention. Patents can also be granted for methods of genetic testing without claiming genes themselves, as illustrated by recent rulings of the European Patent Office Board of Appeal. These rulings upheld patents granted to the biotechnology firm Myriad Genetics and the University of Utah, both based in Salt Lake City, for *BRCA1*-related cancer tests, in the face of considerable opposition from European scientists.

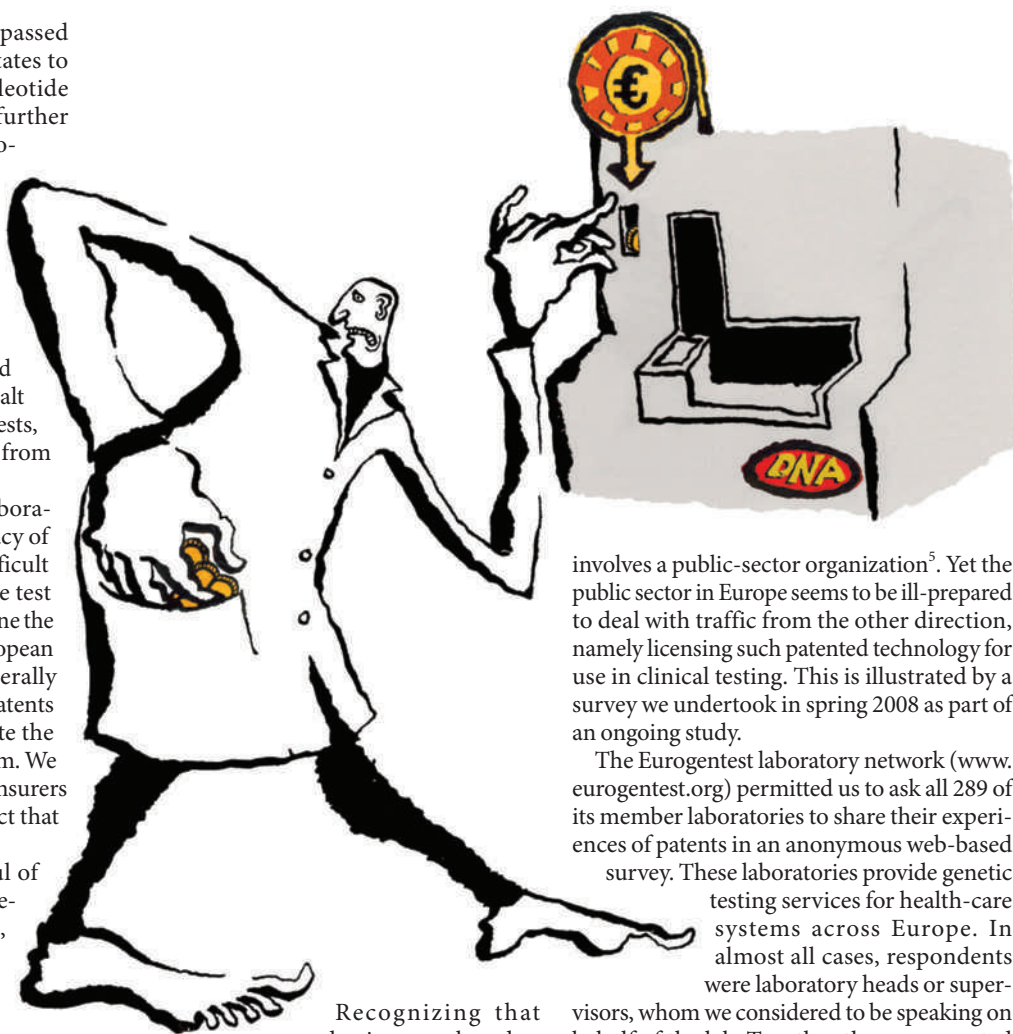
Some European clinical-genetics laboratories refuse to recognize the legitimacy of such patents, arguing that they are difficult to invent around, excessively increase test prices and hinder innovation<sup>1</sup>. To examine the extent of such effects we surveyed European labs and found that, in fact, they generally had little experience of dealing with patents and require more support to negotiate the changing patent landscape around them. We suggest that public and private health insurers will have to come to terms with the fact that costs could rise.

Although we found only a handful of cases in which there was patent enforcement against testing labs in Europe, the financial stakes are rising. More companies are investing in genetic diagnostics, and private market analyses have projected high annual growth in the next 5 years<sup>2</sup>.

## Fair and reasonable

As the market matures, so too must the views of those offering or buying genetic tests. For example, there are different perceptions of what constitutes 'fair' or 'reasonable' licensing fees between non-commercial clinical scientists and biotechnology investors<sup>1</sup>. Such problems are compounded by the licensing strategies of some patent holders, who could deny laboratories the choice to offer a test<sup>3</sup>.

Health-care systems must respond to the possibility that the tests they rely on may be patented by organizations seeking a return on their investments high enough to compensate for the risk of commercial failure.



B. MELLOR

Recognizing that academic research underpins the growth of the biotechnology industry, European countries have strongly promoted patenting in the public sector<sup>4</sup>. This had been spurred on by the lucrative profits of some universities and by the 1980 Bayh-Dole Act in the United States, which permitted patenting of research supported by federal funding. Similarly, the European Parliament's 1998 law (Directive 98/44/EC) that encouraged gene patents is itself a policy response to the perceived economic value of biotechnology patents.

As a result of such stimuli, one in three patent applications on human genetic material

involves a public-sector organization<sup>5</sup>. Yet the public sector in Europe seems to be ill-prepared to deal with traffic from the other direction, namely licensing such patented technology for use in clinical testing. This is illustrated by a survey we undertook in spring 2008 as part of an ongoing study.

The Eurogentest laboratory network ([www.eurogentest.org](http://www.eurogentest.org)) permitted us to ask all 289 of its member laboratories to share their experiences of patents in an anonymous web-based survey. These laboratories provide genetic testing services for health-care systems across Europe. In almost all cases, respondents were laboratory heads or supervisors, whom we considered to be speaking on behalf of the lab. Together they represented the views of 77 labs in the public sector (from hospitals, universities and government labs).

The response rate is lower than that of a similar survey in the United States<sup>6</sup>, perhaps indicating that the issue of patenting is seen as less important in European labs. Indeed, in telephone calls to non-responders,

lack of time was the only reason given for not participating. We detected no response bias in nationality or laboratory size. Because of low response numbers per country, we report on the experiences of European labs overall rather than drawing national-level inferences.

Our findings highlight poor aware-

**"Often labs fail to check for patents, may be able to work around a patent or challenge its validity."**

ness of patent licensing conditions in the European Union. Only 22% of responding labs (17) thought they performed tests on patented genes, while 55% (42) said they did not and 23% (18) did not know. Notably, of the 17 labs testing patented genes, just one lab paid royalties directly, another lab had negotiated with a licensor, but did not pay. Two further labs were aware that royalty fees were included in test-kit prices. There is little evidence of labs being put off patented tests by licensing costs. Only four labs reported costs to be prohibitive, whereas 60% (46) said they did not know enough about licensing costs to comment.

This lack of awareness may be because labs sometimes buy kits for which a patent licence is bundled in the purchase price. In other cases, labs simply develop in-house tests with little concern for patent infringement<sup>3</sup>. These 'home brews' vary in quality but are often less expensive than buying commercial kits<sup>7</sup>. Nevertheless, their use could prompt legal action by patent holders. We did not explicitly ask lab supervisors to state if they infringed patents as this may have discouraged some from responding to the survey. Furthermore, such a question would assume the act of infringement is clear cut. Often labs fail to check for patents, and in other circumstances may be able to work around a patent or challenge its validity.

When it comes to support in dealing with patents, 37% (28 of 75; some did not respond) of public labs reported that they did not have sufficient information or support to deal with patent-related issues. Only 32% (24/75) said they were confident that they did. Furthermore, when asked if the legal advice they needed was available, 26% (19/73) said yes, but for 33% (24/73) legal advice was not available.

### No menace in Europe — yet

For all the concern from geneticists about the effect of patents on genetic diagnostics (see 'The fear of patenting'), there is still relatively little evidence of friction in Europe between patent owners and labs. We found that just 4% (3/77) of responding public-sector labs have ever been prevented from offering a testing service because of a patent-related issue. Three out of six private companies that also responded to our survey said they have been similarly prevented. These figures are still low compared with previous surveys in the United States where 25% (30/122) of labs reported not being able to offer a service as a result of patents held by other organizations<sup>6</sup>. This may be due to a

### The fear of patenting

Comments made on our survey reflect the fears, both real and perceived, surrounding gene patents in the European Union:

**"Enforcement of intellectual-property right would ruin most of the private laboratories. It's a threat."**

**"The costs for paying royalties are not covered by the fees paid by health-insurance companies, nobody is aware of these costs."**

**"The system will collapse, because the patent owners will never get their money, except in the very big labs, or via kits."**

delayed effect, either because patents take much longer to grant in the European Union than in the United States, or because patent owners are yet to take action. But, at this time, our data suggest the concerns about patented genetic inventions have been overstated in Europe.

One possible reason for the trans-Atlantic difference is that fewer genetic inventions are patented in Europe. A 2007 study of human nucleotide sequence patents followed more than 15,000 patent applications filed before December 2003 and found that only 750 had been granted in Europe compared with more than 5,000 in the United States by 2005 (ref. 8). In part, this reflects the higher costs of patenting and the historically higher patentability bar in Europe in this field. Thus, many US inventors

have not patented in Europe to the same extent that they have domestically, even though they have had the opportunity to do so. Other reasons for differences in observed enforcement may be due to differences in patent law, market size and patterns of service delivery that make litigation by patent owners (anywhere

in the world) less attractive in the European Union. In addition, the European Commission monitors closely the implementation of the EU biotechnology directive and its effects<sup>9</sup>.

Although gene patents might not be as menacing as some geneticists have suggested<sup>1,10</sup>, government and health-care systems can no longer afford to ignore patents on genetic inventions. Their current approach might be perceived as hypocritical. Governments continue to promote patenting as a means to improve the return on investment in scientific research, and hospital labs operate increasingly like businesses by charging for their services.

Furthermore, labs generally offer many types of genetic tests, and a growing number use multiple nucleotide sequences, for example, in DNA 'chips' (microarrays). Thus it is increasingly likely that a test that laboratories wish to offer will be affected by a patent at some point. Accordingly, labs are more likely to be called on to respect valid patents, negotiate licences

for expensive technologies and pass costs on to their clients. Equally they must be able to spot invalid patents or workarounds. Governments and health-care systems must ensure patients can access useful tests, provide support to labs facing commercial challenges and pick up the bill. This is the logical consequence of current policy, and the price we pay for the commercialization of science. ■

**Sibylle Gaisser** is at the Fraunhofer Institute for Systems and Innovation Research, 48 Breslauer Street, 76139 Karlsruhe, Germany; **Michael M. Hopkins** is at SPRU: Science and Technology Policy Research, at the University of Sussex, Brighton BN1 9QE, UK; **Kathleen Liddell** is at the Centre for Intellectual Property and Information Law, University of Cambridge, Cambridge CB3 9DZ, UK; and **Eleni Zika** and **Dolores Ibarreta** are at the European Commission Joint Research Center, Institute for Prospective Technological Studies, Edificio Expo, Inca Garcilaso, E-41092 Seville, Spain.  
e-mail: m.m.hopkins@sussex.ac.uk

**"This is the logical consequence of current policy, and the price we pay for the commercialization of science."**

1. Aymé, S., Matthijs, G. & Soini, S. *Eur. J. Hum. Genet.* **16**, S3-S9 (2008).
2. Carlson, B. *Genet. Eng. News* **29**, 14-15 (2009).
3. Matthijs, G. & Hodgson, S. *Clin. Med.* **8**, 58-60 (2008).
4. Lacasa, I. D., Reiss, T. & Senker, J. *Sci. Pub. Policy* **31**, 385-395 (2004).
5. Hopkins, M. M., Mahdi, S., Thomas, S. M. & Patel, P. *The Patenting of Human DNA: Global Trends in the Granting, Filing and Exploitation of DNA Patents. A report for the European Commission Sixth Framework Programme (SPRU, 2006)*. Available at [www.sussex.ac.uk/spru/documents/patgen\\_finalreport.pdf](http://www.sussex.ac.uk/spru/documents/patgen_finalreport.pdf)
6. Cho, M. K., Illangasekare, S., Weaver M. A., Leonard, D. G. B., & Merz J. F. *J. Mol. Diagn.* **5**, 3-8 (2003).
7. Zika, E., Gurwitz, D. & Ibarreta, D. *Pharmacogenetics and Pharmacogenomics: State-of-the-art and Potential Socio-Economic Impacts in the EU* (Institute for Prospective Technological Studies, 2006). Available at <http://ftp.jrc.es/EURdoc/eur22214en.pdf>
8. Hopkins, M. M., Mahdi, S., Patel, P. & Thomas, S. M. *Nature Biotechnol.* **25**, 185-187 (2007).
9. *Development and Implications of Patent Law in the Field of Biotechnology and Genetic Engineering Report from the Commission to the Council and the European Parliament* (2005). Available at <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2005:0312:FIN:EN:PDF>
10. Matthijs, G. *Br. Med. J.* **329**, 1358-1360 (2004).

This article represents the authors' opinions and not those of the European Commission.

See Editorial, page 386, and online at <http://tinyurl.com/dlnksh>.



## BOOKS &amp; ARTS

# New technologies, same old politics

Election campaigns are increasingly being staged online, but digital innovation has brought few new voices into the political debate, explains former parliamentarian **Richard Allan**.

## The Myth of Digital Democracy

by Matthew Hindman

Princeton University Press: 2008.

198 pp. \$22.95, £16.50

Both utopian and dystopian interpretations have been made of the Internet's influence on many spheres of life — and democracy is no exception. Commentators have credited the technology with the power to do everything from destroying representative democracy as we know it, to ushering in a golden age of citizen engagement. Absent from much of this debate is evidence-based analysis of the effects of the Internet on the business of politics. Many theories have been built on nothing more than anecdote, inference and assertion.

In *The Myth of Digital Democracy*, political scientist Matthew Hindman fills important gaps in the evidence base, and does so accessibly. The title might suggest that he argues that the Internet has little influence on politics, but his analysis is much richer than a simple contradiction of the utopian end of the spectrum. Hindman addresses two key questions about digital democracy: what does the Internet do to the day-to-day business of politics, and how does it affect the distribution of powerful political voices?

Digital technologies are changing the business of politics by opening up new avenues for fundraising and communications, and thus altering the infrastructure of political campaigns. Websites allow money to be raised from a broader support base; e-mail and social-networking techniques can be used to organize people around campaign activities. Hindman draws his evidence largely from Howard Dean's high-profile use of the Internet in his attempt to secure the US Democratic Party's nomination for presidential candidate in 2004. Analysis of Barack Obama's recent campaign is included, although the campaign was just getting under way when the book was being written.

Today, digital campaigning is considered effective and essential for any political party's success. This is a far cry from the early 'brochureware' days, when a candidate would derive more electoral value from the traditional media coverage of their website launch than from anything they did online. Although political campaigns still depend on the 'high-street' model of traditional media and local



The Internet opens up avenues for campaigning, but traditional media sources dominate.

infrastructure, evidence confirms that digital technology is useful in supporting this.

Hindman observes, from both the Dean and Obama campaigns, that online activities may be even more significant in a US presidential election than in others because of the primary system. During the early stages of a primary battle, the candidates do not have access to the party's high-street infrastructure and so depend more on digital techniques. Online fundraising, endorsement videos and social-media supporter networks can create a sense of momentum that will in turn be reported in more traditional media.

Most of Hindman's book is directed towards the second, more significant, question of whether digital technologies change the balance of powerful political voices. There is much interest in whether the Internet can empower groups, such as younger people, who are seen as disengaged from the traditional political process. Hindman's answer is in line with the 'myth' of his book title: political voices remain heavily filtered and concentrated on the Internet.

Using data from automated tools that analyse links between websites, Hindman demonstrates that search engines have a powerful effect in concentrating the sites that people visit to find political information. This is because a small number of sites consistently rise to the top of search lists because they have many links from

other sites, and incoming links are used to assign priority by search algorithms. Political influence will be strongest in this handful of heavily linked websites, many of which belong to traditional media organizations. These will therefore continue to be of most interest to politicians.

Website traffic analysis by the company Hitwise reveals that few people visit specialized political sites, indicating that these are less significant than mainstream news sites. Hindman then looks at blogs, showing that a small number of new players with influential political voices have entered the media space. But these are very few; the majority of bloggers have little influence and the successful ones are less diverse than might be assumed.

The book does not address all the potential democratic applications for digital technology, such as viral e-mail marketing and social networking, which are being looked at by other scholars. In each case we need to avoid being dazzled by digital 'magic' and analyse, as Hindman does, whether innovation is really leading to qualitative changes in the political process.

The Internet has already become embedded in politics in a number of countries: in Estonia, for instance, there is a major online component to the whole democratic process, including voting. Many smaller political organizations, including some on the extremist fringes, are innovating with digital technologies to build

J. MOORE/GETTY

campaigns quickly and cheaply. Yet presidential elections in the United States continue to be major landmarks in this area because of their vast campaign budgets, global media attention and the availability of cutting-edge expertise donated by employees of the major US-based Internet companies.

Political parties everywhere have great interest in digital campaigns, especially on the back

of Obama's success; it is now recognized that online activity has moved from an optional extra to an essential element of campaigning. These campaigns may bring different supporters, donors and activists into the political process. We would be right, however, to follow the considered approach of this book in not assuming that enhanced automation of campaigns will effect significant changes in political power.

Based on current evidence, any claims that we are reaching a digitally powered democratic Utopia are indeed more myth than reality. ■

**Richard Allan** is director of global policy and government affairs at Cisco Europe, 9-11 New Square, Bedford Lakes, Feltham, Middlesex TW14 8HA, UK, and was the Member of Parliament for Sheffield Hallam, UK, from 1997 to 2005. e-mail: ricallan@cisco.com

## Science for the greater economic good

### **Tapping the Riches of Science: Universities and the Promise of Economic Growth**

by Roger L. Geiger and Creso M. Sá

Harvard University Press: 2009. 262 pp.

\$39.95, £29.95

When I moved back to Cambridge, UK, in 2002 after a 21-year absence, the biggest difference I noticed was the high house prices in the city centre. These reflected the high earnings, not of senior academics but of businessmen, some millionaires, working in the surrounding science parks. The university had changed little, but spin-out companies, entrepreneurialism and wealth generation brought real-world concerns closer to the ivory tower. A 2006 study (see <http://tinyurl.com/bv8xk6>) concluded that if the university disappeared, 77,000 local jobs and a net value in the region of £21 billion (US\$29.5 billion) would go with it.

*Tapping the Riches of Science* looks at how universities in the United States have similarly become important generators of local and national economic growth. The Bayh–Dole Act of 1980 gave universities the ownership of intellectual property generated by federal research grants, and the responsibility for exploiting that intellectual property. Such roles have been largely viewed with suspicion. Derek Bok, former president of Harvard University, wrote in *Universities in the Marketplace* (Princeton University Press, 2003) that universities were becoming the handmaidens of football leagues, pharmaceutical industries and online providers of education. In *Science For Sale* (University of Chicago Press, 2007), Daniel Greenberg stressed the conflicts of interest between faculty and the pharmaceutical and biomedical industries.

With its dispassionate analysis, *Tapping the Riches of Science* is welcome. It includes the upsides and downsides of the rise of economic development, which has become the fourth mission of US universities after teaching, research and outreach. Authors Roger Geiger and Creso Sá emphasize that universities are in the driving seat and stand to benefit most



SCIENCE RESEARCH, UNIV. CAMBRIDGE

**Innovations by university researchers, such as this rapid visual test for HIV, are crucial to economic growth.**

from links to industry: “The universities of the twenty-first century have essentially exploited the opportunities inherent in economic relevance to garner increased resources from both industry and government.”

Geiger and Sá point out that university research has long driven economic growth in the United States. After the Hatch Act of 1887, federal funds allocated to land-grant colleges and universities established agricultural research centres. As a university mission, technology transfer became entrenched after 1945, with defence-related labs being formed at the Massachusetts Institute of Technology and the University of California, Berkeley, among others. Intriguingly, it is the campuses that gained experience of working with industry and the military that embraced the economic agenda earliest; Yale University and Harvard University were among the last. Health science also enjoyed rapid expansion in parallel with the growth of the US National Institutes of Health.

In recent years, universities have come to be viewed as small but vital players in the national innovation system. The twin paths of innovation

— the support of large corporations through licensing agreements, and the incubation of new industries through spin-outs and start-ups — have provided challenges in their interactions on campus, including questions over ownership of intellectual property, conflicts of interest and debates about subsidized access to labs. Many early partnerships were the result of enterprising senior individuals, but as these linkages have grown, institution-wide adjustments have become necessary to accommodate them.

The idea that universities will ‘get rich quick’ through the exploitation of their own intellectual property has been replaced by a more realistic view of the ‘public good’ element of local economic growth. The dichotomy between public and private, between seeking profit and crystallizing the value of what was originally public-good research will remain a source of debate, but we are now over the naive hurdle that the two cannot mix.

The relationship between business and the academic core is fascinating, affecting the development of disciplines and the appointment of faculty. The formation and governance



of specialized multidisciplinary institutes, both within the university and with government or corporate partners, is described in detail.

US universities with a strong science and engineering base have gone the farthest down this route. This trend is reflected in the book, which somewhat under-represents the role of arts and humanities faculty in creating partnerships and innovation. The pace of innovation may be slowed in universities that are governed from the bottom-up by faculty, whereas it is more likely to be pushed by governing bodies that have a majority of external members.

*Tapping the Riches of Science* focuses on the large and diverse US university system. I would like to see this study extended to consider the effects of globalization. Not all university systems are as well placed to be a partner in global consortia. For example, entrepreneurial faculty may fare better in the independent universities of the United Kingdom compared with those in Japan or many countries in Europe, where

professors are essentially civil servants.

Moreover, it is the global problems that are coming to the fore — overpopulation, poverty, climate change, financial instability and inequity. In 20 years, I anticipate that a similar book will describe how universities in different global consortia will have succeeded in getting their institutional arms around these complex problems. Providing global solutions will involve a greater proportion of the faculty, not just from science and engineering, necessarily working in multidisciplinary teams. Such collaboration will further erode Cardinal Newman's nineteenth-century vision of a university as a self-governing community of disinterested scholars. ■

**Michael Kelly** is chief scientific adviser to the UK Department for Communities and Local Government, and the Prince Philip Professor of Technology in the Department of Engineering, University of Cambridge, Cambridge CB3 0FA, UK. e-mail: mjk1@cam.ac.uk

emphasizes, is that systems as diverse as cells, economies and ecosystems, as well as the human brain, all process information, and do so in a way that makes them rich, adaptable and hard to understand. The book hits its stride in its latter half, with an insightful survey of recent developments in complex-network theory and scaling in biology.

Especially valuable is the book's exploration of recent attempts to categorize the dynamics of cellular automata — simple systems that act as models for the study of rich dynamics. Some of this work, under the name of computational mechanics and linked to the ideas of Mitchell's former colleague, the late Jim Crutchfield, probes the fundamental 'information physics' of complex systems in general. This focus of the book is commendable, as much of the literature of complex-systems research dwells on more expansive philosophical themes at the expense of the 'boring' details of specific models. Yet intense scrutiny of such models may ultimately reveal clues to solving Bedau's mystery.

Mitchell touches on the many practical applications of this science, ideas put into practice by forward-looking companies such as Cisco and Capital One. The book is timely, given that many analyses of the present financial crisis have concluded that the key issue is how markets have outstripped our ability to understand them.

It has become fashionable in recent years to criticize complex-systems science for generating too much hype and not offering enough practical insight. But insights into truly complex problems do not come easy. Mitchell's welcome book makes it clear that this field is making steady, if slow, progress. ■

**Mark Buchanan** is a writer based in the United Kingdom and author of *The Social Atom*. e-mail: buchanan.mark@gmail.com

## Bringing clarity to complexity

### **Complexity: A Guided Tour**

by Melanie Mitchell

Oxford University Press: 2009.

368 pp. £14.99, \$29.95

For several decades, scientists studying complex systems — rich, collective systems such as ant colonies, economies and cells — have spoken of 'emergence', the mysterious process by which the collective whole acquires resilience, adaptability and other surprising properties, even though its components are simple. The archetypal example is the ant colony, which manages to forage intelligently for food and organize collective defence by exploiting the limited skills of its individual ant citizens.

The ideas of complexity have spread across science, and emergence has become a buzzword. Philosopher Mark Bedau has suggested, however, that it poses a puzzle, as it demands that two seemingly contradictory statements must be true. In complex systems, organized phenomena at higher levels depend on processes at lower levels: everything in a cell, for example, depends on the processes of atomic physics. Yet phenomena emerging at higher levels gain autonomy from lower levels: the body's organs and their interactions can be described and explained without reference to atomic physics.

How can something be dependent and autonomous at the same time? And why do so many systems in nature show this hierarchical

organization? No one has answered these questions, but in *Complexity*, computer scientist Melanie Mitchell of the Santa Fe Institute, New Mexico, offers a valuable snapshot of the growing field of complex-systems science from which the answers may eventually arise.

Mitchell explores the historical roots of this area in the work of visionaries such as Henri Poincaré and Edward Lorenz in dynamical-systems theory, and of John von Neumann, Alan Turing and others in computation. The unifying feature of complex systems, Mitchell



As a colony, ants perform complex tasks that individuals could not achieve alone, such as tending larvae.

MEUL/ARCO/NATURE PICTURE LIBRARY



# Scientific symbolism

In his dynamic 1891 ceiling decoration for Paris's city hall, Paul-Albert Besnard depicts the unveiling of truth by the sciences and makes a statement as strong as that of the Impressionists, explains **Martin Kemp**.

Politics and big science invariably mix — in the eyes of those in power, at least. Nowhere was this more true than in late-nineteenth-century France, where the radical republican government sought to identify with the rational truths of science, the prestige of national scientific discoveries and the dynamics of technological progress. Large-scale art in state buildings was a major vehicle for the promotion of this identity.

The artist who succeeded most effectively in developing a new mode of painting to express this ideal was Paul-Albert Besnard (1849–1934). Now little noticed in the standard histories of French painting, he was regarded as a giant of art in his day, and was accorded the unusual accolade of a state funeral. He worked on large murals and ceiling paintings in Paris — at the Sorbonne, the School of Pharmacy and at the Hôtel de Ville (city hall).

On the ceiling of the Salle des Sciences in the Hôtel de Ville, he painted an allegory that shows how Truth, pulling the sciences in her train, pours her light on mankind. Completed in 1891, it departs radically from traditional depictions of the heavens, which were populated by gods, goddesses and symbolic figures on banks of strategically placed clouds.

Two great orbs, incompletely seen, dominate Besnard's composition. The radiant orb is the Moon, with its craters and 'seas', and the darker one is Earth. The space beyond is dense with stars and galaxies, dizzily swirling in a cosmic vortex. Truth, personified as an evanescent woman dressed in white, drags a gloomy veil — signifying obscurity — away from the sciences in her train, each crowned with laurels. The most prominent of them metaphorically points our way towards enlightenment.

Beside Truth is the phosphorescent figure of Light, ecstatically bearing a flaming sheaf of pure illumination for the benefit of mankind.

A snaking parade of humans emerges from the gloom, dragged from their primitive state by what science can disclose about the world and the Universe.

All this is remote from current taste in the visual arts. We prefer to look to the Impressionists and post-Impressionists of this period; if we want an image of the heavens in painting,

knowledge, but also a moral doctrine, directly opposed to religious doctrine."

Mauclair's rapturous account of the Hôtel de Ville's ceiling aims to evoke a kind of ecstasy of rational wonder, in which the sciences replace the saints as a subjects of devotion. "Here," he exclaims, "is a hymn to fire, principle of life and of the energies of thought and understanding."

The conflagration of colours in Besnard's ceiling decoration allies the brilliance of Impressionist colour with academic draftsmanship and a dynamic of geometrical composition, in a way that Mauclair saw as ushering in the true art of the new century.

What Besnard is not doing is illustrating scientific facts. It is true that we can recognize the Moon crater Tycho, with its prominent rays, close to the contour of Earth. Large photographs of the Moon had become popular exhibits at the various universal exhibitions, not least that in Paris in 1867. And he seems to be drawing on hazy images of the nebulae. But accuracy in the literal sense is not his goal.

Besnard is creating an allegory of the spirit of science — indeed, of its spiritual dimension — one that was fitting for a new age of scientific enlightenment, in which humanity

moves ever onwards and upwards towards true understanding of the cosmic mysteries. He is saying that this is the highest goal for mankind, and that a new art is needed for the brave new age.

The future for art was not to lie predominantly with Besnard. Such grandiose decorative schemes fell out of favour, as did his manner of painting, for a variety of reasons. His themes, however, as explicated by Mauclair, are still with us — science and religion, science as enlightenment and science as redeemer. ■

**Martin Kemp** is emeritus professor in history of art at the University of Oxford, Oxford, UK.



Paul-Albert Besnard's mural reflected the new age of scientific enlightenment.

we are likely to describe Vincent van Gogh's *The Starry Night* (1889). But Besnard's depiction is highly original. He was striving to forge a new kind of *symbolisme scientifique* that would sweep away the old forms of classical and religious allegory.

The term 'scientific symbolism' comes from Camille Mauclair's 1914 monograph on Besnard. Mauclair defines it as "the plastic and coloristic incarnation of scientific notions in decorative allegories". Poet, novelist, travel writer and critic, he may be taken as a spokesman for Besnard. The new form of allegory eulogizes science "not only as experimental

MAIRIE DU 1<sup>ER</sup> ARRONDISSEMENT, PARIS/LAURENCE GIRAUDON/BRIDGEMAN ART LIBRARY

## NEWS &amp; VIEWS

## PALAEOLOGY

## Beyond the Age of Fishes

Michael I. Coates

**Discovery of an unusually intact and ancient fossil fish provides further evidence that the search for modern vertebrate origins requires breaking out of the Devonian and into the preceding period.**

As a rule, the earliest fossils of living groups tend to be scrappy, and such fragments lend themselves to contentious interpretations. For 'bony fishes', Osteichthyes — the division of vertebrates that includes everything from humans to halibut — the record of articulated fossils peters out within the Lower Devonian<sup>1</sup>, some 400 million years ago. Earlier stretches of osteichthyan history are littered with fossil detritus, such as isolated teeth and scales. In certain instances, bits and pieces have been reassembled into conjectural species<sup>2–4</sup>, some of which have surprising combinations of anatomical features<sup>5</sup>. On page 469 of this issue, Zhu *et al.*<sup>5</sup> introduce a fresh — albeit long-dead — fish into this poorly resolved patch of vertebrate evolution. Crucially, this piscine offshoot of our own distant past is both unusually intact and exceptionally old.

So what kind of fish is it? A summary of vertebrate diversity helps to make sense of the answer. Of the 51,000 or more living species of vertebrates, 99.9% have jaws: these are the gnathostomes. Gnathostomes include the bony Osteichthyes and the cartilaginous Chondrichthyes. Chondrichthyes (sharks, rays and chimaeras) account for only 2% of gnathostome species, the Osteichthyes accounting for the other 98%. Around half of the Osteichthyes are Actinopterygii, or 'ray-finned fishes', and half are Sarcopterygii, or 'lobe-finned fishes'. Actinopterygians include some 28,000 species, from zebrafish to bichirs, and living sarcopterygian fishes are limited to three genera of lungfishes and one coelacanth. Land-dwelling tetrapods constitute the remaining majority of sarcopterygians.

Thus far, the origins of these major divisions of today's gnathostomes can be traced back to the Devonian, between 416 million and 359 million years ago, the Age of Fishes. Fossils that are clearly chondrichthyan are known from around 405 million to 400 million years ago<sup>6</sup>, but we have little idea as to whether these belong within the living radiation, the 'crown group', or represent side branches of their common



**Figure 1 | Newcomer to the Silurian seascape.** This classic view of Silurian marine life, published in the 1940s, is rich in invertebrates (corals, molluscs, arthropods, echinoderms, and more besides). But it lacks fish. Armoured jawless fishes existed throughout the Silurian (443 million to 416 million years ago), alongside early jawed fishes (placoderms and acanthodians, extinct groups whose affinities are the subject of debate<sup>8,10</sup>). A representative of modern fishes, *Guiyu oneiros*<sup>5</sup> (inset), can now be added to the picture. *Guiyu* is a Silurian-aged member of the sarcopterygians (extant representatives of which include lungfishes, the coelacanth and all tetrapods). What else might be absent? Evidence of early actinopterygians (ray-finned fishes) and chondrichthyans (sharks and chimaeras) must be lurking out there, somewhere in the Silurian sediments. (Silurian scene by Z. Burian. Fish reconstruction by B. Choo.)

ancestry, the 'stem group'. As for osteichthyans, although it is agreed that fossils from the earliest Devonian<sup>2,7</sup> belong within the crown, osteichthyan fragments of less-certain affinity are also known from the Late Silurian<sup>3</sup>, 423 million to 416 million years ago.

But there's more to this story, because the question of gnathostome origins also involves a pair of extinct groups of gnathostomes known to appear earlier in the geological record, the placoderms and acanthodians<sup>1</sup>. Importantly, recent analyses<sup>8</sup> have begun to reveal new relationships between early vertebrates, in which acanthodians and placoderms are scattered among the early divisions of gnathostome evolution; acanthodians, in particular, are cropping up on chondrichthyan and osteichthyan stem groups. The straightforward message

is that the origin of modern gnathostomes is not a Devonian phenomenon, after all. The basal divergence between osteichthyans and chondrichthyans occurred somewhat earlier.

This, then, is the context within which to place *Guiyu oneiros*, the new species of early osteichthyan named and described by Zhu *et al.*<sup>5</sup>. Preserved in 418-million-year-old limestone in what is now southern China, the fossils of *Guiyu* show the skeletal anatomy of a small sarcopterygian, around 33 centimetres long. The very fact that *Guiyu* can be identified as a sarcopterygian provides further and arguably clinching evidence that a whole series of major branching events within the gnathostome crown group must have taken place well before the end of the Silurian.

Like any other fossil, *Guiyu* has a mixture of



primitive and advanced features. With regard to its anatomical completeness, *Guiyu* provides exceptional corroboration for the decidedly odd reconstruction of the early osteichthyan genus, *Psarolepis*<sup>2</sup>. Cobbled together from a disparate set of fossils, the incongruent suite of features<sup>9</sup> displayed by *Psarolepis* has been viewed with caution. Now, it turns out to be thoroughly plausible. Like *Psarolepis* and other sarcopterygian fishes (including *Latimeria*, the living coelacanth), the braincase of *Guiyu* is divided into separate front and rear units. Like *Psarolepis*, the cheek bones resemble those of early actinopterygians. Like *Psarolepis* and many other early gnathostomes<sup>1</sup>, including at least one chondrichthyan<sup>6</sup>, the shoulder girdle bears a spine in front of the pectoral fin. Similarly, the dorsal-fin spine and anterior spine-bearing plate of *Guiyu* are probably primitive. These are all widespread features of early gnathostomes, and seeing such characteristics in *Guiyu* provides a first glimpse of the sequential order of anatomical changes that resulted in the standard set of sarcopterygian traits.

The evolutionary tree proposed by Zhu *et al.*<sup>5</sup> (see Fig. 5 on page 473) adds to a growing set of analyses of early osteichthyan and gnathostome interrelationships<sup>8,10</sup>. Uncertainties still surround the branching pattern of non-osteichthyans, but the addition of *Guiyu* to the cast of early fishes does not change the basic pattern of interrelationships among early osteichthyans. Instead, it adds support to notable consistencies in the emerging pattern of sarcopterygian evolution, including the clustering of some of the earliest-known examples to form an as-yet unnamed group.

Finally, what does the conclusion that *Guiyu* is unequivocally sarcopterygian imply? On the whole, early fossils are thought to be unreliable as minimum-date markers of evolutionary branching events<sup>11</sup>, because they are less complete and/or lack the full anatomical signature of the group to which they are assigned. *Guiyu* might be an exception that proves the rule, for it provides a new and exceptionally reliable earliest fossil marker for a major split in vertebrate evolution. By pushing a whole series of branching points in gnathostome evolution out of the Devonian and into the Silurian, the discovery of *Guiyu* also signals that a significant part of early vertebrate evolution is unknown (Fig. 1).

The new shape of the gnathostome tree shows that early sarcopterygians, as well as actinopterygians and chondrichthyans, ought to be turning up in Silurian sediments. But where are they? Modern fish groups have Silurian roots, but these fishes are consistently absent from existing scenarios of Silurian life. The discovery of *Guiyu* should provoke a rash of new fieldwork and a fresh look at existing collections of pre-Devonian fossils. ■

Michael I. Coates is in the Department of Organismal Biology and Anatomy, University of Chicago, Chicago, Illinois 60637, USA.  
e-mail: mcoates@uchicago.edu

1. Janvier, P. *Early Vertebrates* (Oxford Univ. Press, 1996).
2. Zhu, M. *et al. Nature* **397**, 607–610 (1999).
3. Botella, H. *et al. Nature* **448**, 583–586 (2007).
4. Basden, A. M. & Young, G. C. J. *Vert. Paleontol.* **21**, 754–766 (2001).
5. Zhu, M. *et al. Nature* **458**, 469–474 (2009).
6. Miller, R. F. *et al. Nature* **425**, 501–504 (2003).

7. Zhu, M. *et al. Nature* **441**, 77–80 (2006).
8. Brazeau, M. D. *Nature* **457**, 305–308 (2009).
9. Ahlberg, P. E. *Nature* **397**, 564–565 (1999).
10. Friedman, M. J. *Syst. Palaeontol.* **5**, 289–343 (2007).
11. Donoghue, P. C. J. & Benton, M. J. *Trends Ecol. Evol.* **22**, 424–431 (2007).

## ASTROPHYSICS

# Quiet is the new loud

Daniel Proga

**Understanding the mechanisms by which matter flows into black-hole systems is pivotal to elucidating how such systems work. It seems that a 'quiet' mass outflow can play a hitherto-unknown part in the process.**

All black-hole systems that accrete matter, regardless of their size, are believed to have very similar components and to operate in a very similar way. Quasars — galaxies with extremely bright nuclei powered by the accretion of matter onto a supermassive black hole — are an example of accreting black-hole systems at one end of the size range. At the other end are their much smaller cousins, black-hole binaries (a black hole with a star companion). Black-hole binaries have fascinated astronomers for years because they go through a cycle of many different activity states. For example, they can be in a state of high accretion and high luminosity, in which they strongly emit both 'soft' (low energy) and 'hard' (high energy) X-rays — the bright/soft state. Another state is one of low accretion and low luminosity, in which the hard-X-ray emission exceeds that of soft X-rays — the faint/hard state.

On page 481 of this issue, Neilsen and Lee<sup>1</sup> report observations of a microquasar — a black-hole binary that has a radio-emitting jet of gas — known as GRS 1915 + 105. This black hole has a mass 14 times that of the Sun, and is accreting gas from its star. The authors have discovered that, as the system changes from the faint/hard accretion state to the bright/soft state, the high-speed jet is replaced by a much slower, X-ray-absorbing wind (Fig. 1). The authors conclude that the wind being launched from the outer regions of the black hole's accretion disk competes with the jet for matter and wins, taking matter away from the disk and halting its flow into the jet.

Many quasars are found only in the bright/soft accretion state. This is not because they do not vary, but rather because they evolve more slowly than their smaller counterparts. Therefore, to understand quasars, astronomers tend to study microquasars, which evolve on timescales that are six to eight orders of magnitude shorter than quasars.

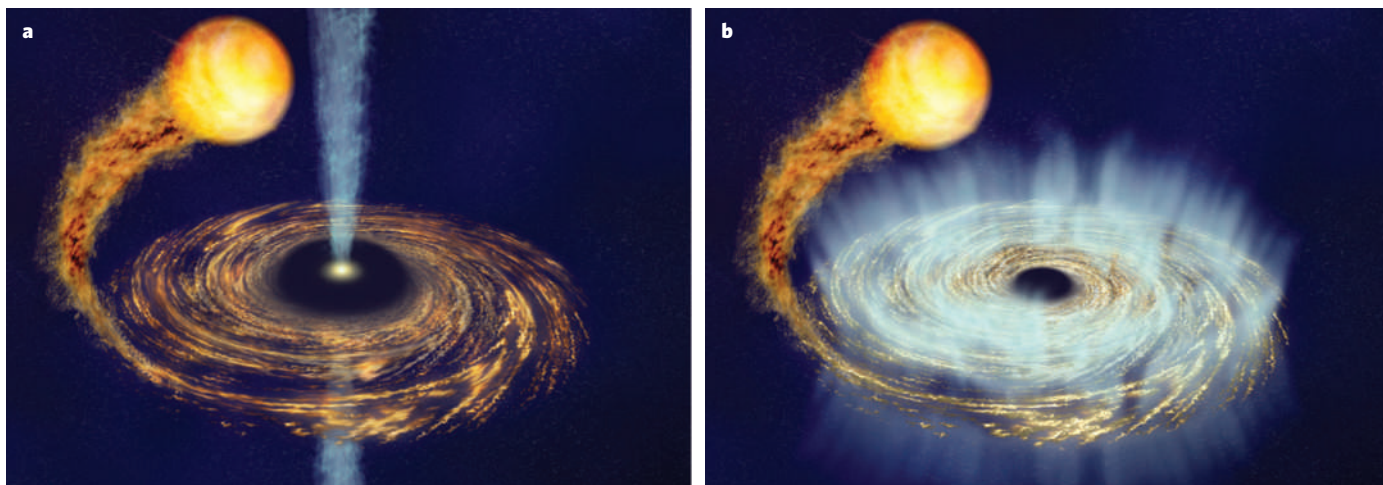
Extensive observations of microquasars have revealed that they can emit radiation over a wide range of energies, from radio to X-rays and  $\gamma$ -rays. Modelling of the observed spectra

shows that this broad emission is possible because microquasars have different sources of energy: ultraviolet and soft X-rays are emitted by the accretion disk; hard X-rays are produced in the disk's corona of very hot plasma; and radio emission is generated by the narrow jet of magnetized plasma<sup>2,3</sup>. Observed changes in emission can thus be linked to changes in the sources of radiation.

Most studies of microquasars have focused on their main, 'loud' components — that is, the disk, corona and jet. These studies have identified a variety of accretion states<sup>4</sup> (up to 14 states in GRS 1915 + 105), and several correlations and anti-correlations among the emissions in different energy bands. For example, the radio emission correlates with X-ray emission in the faint/hard state<sup>5</sup>. In addition, such studies have yielded clues to one of the main mysteries: how matter that plunges onto the black hole can avoid crossing its event horizon (the boundary below which nothing can escape), and instead can escape as a jet with velocities close to the speed of light. One clue is that jets occur when matter accretion is in the form of a thick flow.

But despite many successes, the most crucial questions, such as how accretion disks work and how jets are produced, remain unanswered. We do not know exactly how the rotational energy of the disk is dissipated, converted to heat and finally radiated away. However, we do have a good physical model for the outward transport of angular momentum in the disk if magnetic fields are present<sup>6</sup>. This is extremely important, because such transport is crucial for accretion to occur in the first place. As for jet production, we still do not understand whether the jets are powered by the rotational energy of the accretion flow or by the energy of the fast-spinning black hole<sup>7,8</sup>. In both cases, magnetic fields are involved in transferring energy to the jet.

Microquasars thus continue to be the subject of intensive observational and theoretical research. Neilsen and Lee's work<sup>1</sup> shows that much can be learned from investigating the 'quiet' components of the systems, such as disk



**Figure 1 | Broad wind takes over.** Black-hole binaries, which consist of a black hole that is accreting matter from a companion star (top left), can be found in many activity states. For example: **a**, a state of low matter accretion and low luminosity; **b**, a state of high accretion and high luminosity. Neilsen and Lee<sup>1</sup>

studied a black hole 14 times as massive as the Sun that is accreting gas from a star of almost the Sun's mass. They found that, as the system changes to its high-accretion state, a broad wind of matter takes control of the accretion process and halts the flow of matter into the narrower, fast-moving jet of gas seen in **a**.

winds. Although these winds can be a source of emission, they mainly show up as absorbers of the continuum emission produced by other components. This absorption is typically limited to spectral lines and is 'Doppler-shifted' towards shorter wavelengths. Hence, we know that the absorbing material (the wind) is moving away from the source of the continuum emission — the accretion disk — and towards an observer. Because the wind is launched from the disk, as opposed to being launched from the companion star, this property can be used to connect the physics of winds with that of accretion disks. For example, Miller *et al.*<sup>9</sup> have argued that the wind in another black-hole accreting system (GRO J1655–40) is driven by magnetic fields in the disk.

By contrast, magnetic fields are not required to account for the wind observed in Neilsen and Lee's system. The authors contend that the energy of the radiation emitted by the innermost part of the disk is high enough to heat up and expand its outer part, thereby producing the wind. Such thermal driving is probably assisted by the momentum of the same radiation, which can give it an extra 'push'.

Neilsen and Lee's study thus raises several questions. Can winds in microquasars be driven in two different ways? How are the jet and wind related? Is the jet a narrow version of the wind? According to the authors, the answer to the last question is no. They argue that the wind and jet compete for the same material. However, they do not suggest a mechanism that might underlie such competition. It is possible that some kind of magnetic-field effect could determine whether a jet or a wind forms, because magnetic fields have been shown to be implicated in the accretion process and in powering jets. But why do magnetic fields produce winds in only some instances? Perhaps the authors' interpretation that radiation — not magnetic processes — is responsible for the winds is correct.

The work<sup>1</sup> also has major implications for jet

collimation in all black-hole accreting systems. Internal processes in the jet are not efficient enough to produce collimation. Some researchers have therefore suggested that the jets are narrow as a result of an external medium, with the disk wind being a strong candidate. But Neilsen and Lee's study contradicts that view, in that the wind seems to suppress — not collimate — the jet.

Their observation is puzzling, and the hope is that simultaneous multi-wavelength observations will elucidate the intricate inner workings of these systems. For example, it would be interesting to know what happens when the wind weakens, and whether the disappearance of the jet. Meanwhile, the lesson we have learned from Neilsen and Lee's study is that, ironically, we can gain insights about the 'loudest' source

of radiation (the accretion disk) by investigating the 'quiet' wind that muffles the source. ■ Daniel Proga is in the Department of Physics and Astronomy, University of Nevada, Las Vegas, 4505 South Maryland Parkway, Las Vegas, Nevada 89154-4002, USA. e-mail: dproga@physics.unlv.edu

1. Neilsen, J. & Lee, J. C. *Nature* **458**, 481–484 (2009).
2. Fender, R. & Belloni, T. *Annu. Rev. Astron. Astrophys.* **42**, 317–364 (2004).
3. Done, C., Gierlinski, M. & Kubota, A. *Astron. Astrophys. Rev.* **15**, 1–66 (2007).
4. Klein-Wolt, M. *et al. Mon. Not. R. Astron. Soc.* **331**, 745–764 (2002).
5. Gallo, E., Fender, R. P. & Pooley, G. G. *Mon. Not. R. Astron. Soc.* **344**, 60–72 (2003).
6. Balbus, S. A. & Hawley, J. F. *Rev. Mod. Phys.* **70**, 1–53 (1998).
7. Blandford, R. D. & Payne, D. G. *Mon. Not. R. Astron. Soc.* **199**, 883–903 (1982).
8. Blandford, R. D. & Znajek, R. L. *Mon. Not. R. Astron. Soc.* **179**, 433–456 (1977).
9. Miller, J. M. *et al. Nature* **441**, 953–955 (2006).

## MOTOR-NEURON DISEASE

# Rogue gene in the family

Kristel Slegers and Christine Van Broeckhoven

**Various gene mutations contribute to the motor-neuron disease known as amyotrophic lateral sclerosis. Further mutations that have been identified could help to clarify the neurodegenerative mechanism in this disorder.**

At a time when it is not uncommon for genetic studies of human conditions to involve tens of thousands of individuals, and more than a million genomic variants associated with marginally higher or lower risks, studies involving just a handful of patients can still bring about a breakthrough. A testimony to this are two studies<sup>1,2</sup> published in *Science*, which report that the heritable form of the disease amyotrophic lateral sclerosis can be caused by mutations in the *FUS* gene.

Amyotrophic lateral sclerosis (ALS) — also known as Lou Gehrig's disease, after the American baseball legend — is an incurable, severely disabling condition in which degeneration of motor neurons causes progressive muscle weakness and spasticity; most patients die of respiratory failure within five years of disease onset. Inherited cases of ALS account for some 10% of patients and are invaluable in the hunt for genes underlying this disorder. Indeed, a few genes have already been implicated in familial





## 50 YEARS AGO

Considerable data are now available on the radiosensitivity of cultivated plants, but very little is known about the tolerance of wild species. We have investigated the doses of  $\gamma$ -radiation needed to prevent weed seeds from growing ... The results obtained indicate that arable soils need treatment with at least 100,000 rads of  $\gamma$ -radiation to inhibit weed growth effectively. This treatment would sterilize any insects, or nematodes present in the soil, and would probably destroy a high percentage of the bacteria and fungi.

From *Nature* 28 March 1959.

## 100 YEARS AGO

Much interest has been aroused in Sussex by the discovery of the greater part of a skeleton of a mammoth (*Elephas primigenius*) on the shore of Selsey Bill. The remains were found below high-water mark in the estuarine or fresh-water deposit of black clay, which underlies the raised beach and coombe rock on that part of the Sussex coast. The thick mass of shingle, which usually covers this deposit, was temporarily removed during the recent stormy weather, and the teeth and broken bones were found projecting from the clay. Probably the whole skeleton was originally present, but when found the bones were already much eroded, and they were scattered over an area about 30 feet square. Both upper and lower molar teeth were recovered, and their condition shows that the animal was immature and of small size. Fragmentary remains, both of the mammoth and of *Elephas antiquus*, have been found at various times in the same deposit in Bracklesham Bay, some of these specimens being now in the British Museum. Indications of complete skeletons are rare. They seem to have been recorded only twice in England, the first in the brick-earth of Ilford, Essex, the second in a corresponding deposit at Ealing, Middlesex.

From *Nature* 25 March 1909.

cases<sup>3</sup>. For instance, the *SOD1* gene is mutated in ALS with relatively high frequency<sup>4</sup>; so is the *TARDBP* gene, which encodes the TDP-43 protein, a frequent constituent of the disease-associated cytoplasmic inclusion bodies found in patients' motor neurons<sup>5</sup>. Furthermore, a link between several genomic regions and ALS has been suggested, one such region being on chromosome 16, where *FUS* is located<sup>6,7</sup>.

Members of a family originating from a Cape Verde island<sup>1</sup> and a British family<sup>2</sup>, with evidence of a genetic cause of the disease on chromosome 16, were instrumental in the identification of *FUS* as an ALS-linked gene. Studying the descendants of first-cousin Cape Verdean grandparents, Kwiatkowski *et al.*<sup>1</sup> identified a 4-megabase-pair region in the ALS-linked locus on chromosome 16 in which both DNA strands were identical by descent in patients but not in healthy siblings. Through subsequent sequencing of the genes in this region, the authors pinpointed a single-nucleotide (missense) mutation in *FUS*. Vance *et al.*<sup>2</sup> independently identified a different missense *FUS* mutation in the British ALS family. What's more, the two teams<sup>1,2</sup> screened other patients with familial ALS and identified 12 additional *FUS* missense mutations in 24 families.

Both teams confirm the pathogenic effect of *FUS* mutations through tissue autopsy of patients who carried the mutation, showing that the motor neurons of these patients contain cytoplasmic inclusions of FUS protein. Further studies in cultured cells expressing *FUS* mutants confirmed the relocation of this protein from the nucleus to the cytoplasm<sup>1,2</sup>. It is noteworthy that, in the autopsy samples, the typical TDP-43 inclusions were absent<sup>2</sup>, indicating that *FUS*-associated ALS has a neuropathological mechanism distinct from that of the TDP-43-associated disease.

Most of the *FUS* mutations identified involved the substitution of evolutionarily conserved arginine amino-acid residues in the carboxy terminus of FUS with other amino acids, emphasizing the crucial role of the arginine residues in the normal functioning of the protein. The age at onset and disease duration, however, varied between carriers of the same or similar *FUS* mutations. For example, Kwiatkowski *et al.* found that in patients who carried a substitution of arginine at position 521 for a glycine residue, the average age of onset was 37.5 years in one family and 60.7 years in another. This observation raises the possibility that other genetic sequence variants in *FUS*, or in other genes, influence the onset and progression of the disease. Such modifying genetic factors could be instrumental in designing drugs that prevent or delay the development of ALS, as they reveal molecular mechanisms that can accelerate or delay the disease's progress.

The identification of *FUS* mutations in patients with familial ALS ends a long period of intense hunting for the genetic defect on chromosome 16 associated with this condition, and provides an impetus for explorative research

into the role of FUS in the pathogenesis of the disease. The question now is how FUS mutations affect motor neurons.

Being predominantly a nuclear protein, FUS is an RNA- and DNA-binding molecule that is involved in many cellular processes, including DNA repair, RNA splicing and transcription, and transport of messenger RNA from the nucleus to other destinations — such as the dendritic processes of neurons — for local protein synthesis at the synaptic junctions between neurons<sup>8</sup>. FUS mutations could lead to a loss of any of its normal functions or gain of toxic functions in the nucleus, resulting in impaired RNA processing. Cytoplasmic aggregates of FUS, for instance, could be toxic to the cell, or lead to sequestration of other proteins in the cytoplasm.

In addition, FUS shows intriguing functional similarities to TDP-43, which is also an RNA- and DNA-binding nuclear protein, and this suggests that the two proteins might have shared RNA targets that affect motor-neuron survival. For both FUS and TDP-43, ALS-causing missense mutations predominantly affect the proteins' carboxy terminus, which contains motifs for RNA processing. The common RNA targets of these proteins could include precursors to motor proteins, cytoskeletal proteins and growth factors crucial for the normal functioning of motor neurons. Intriguingly, TDP-43 pathology is associated with loss of the growth factor progranulin in neurodegeneration<sup>9</sup>, whereas one target of FUS is insulin-like growth factor-1 (ref. 10). Furthermore, FUS and TDP-43 are not the only RNA-processing proteins implicated in ALS and other motor-neuron disorders<sup>11</sup>, underscoring the fact that abnormal RNA metabolism may represent a central pathway for motor-neuron degeneration. So genetic screening of other genes involved in RNA metabolism — such as *ELP3*, which has recently been associated<sup>11</sup> with ALS — is warranted.

Most of the *FUS* missense mutations reported in the two papers<sup>1,2</sup> fit an autosomal-dominant inheritance pattern — that is, mutation in a single copy of the gene is sufficient to cause disease — and an age-dependent increase in the proportion of individuals developing the disorder. The mutation detected in the Cape Verdean family, however, caused ALS only in those who carried it in both copies of their *FUS* gene (homozygous), and so this conforms to a recessive pattern of inheritance<sup>1</sup>.

Cultured cells expressing a single copy of the Cape Verdean mutation did not show the same pattern of cytoplasmic FUS retention as that observed for the dominant mutations; but, compared with normal cells, they did show a slight, albeit significant, alteration in the cellular localization of this protein. This mild *in vitro* effect of the single-copy Cape Verdean mutation might reflect a milder or related clinical expression of ALS in individuals carrying this mutation in a single copy of their *FUS* gene, although this possibility awaits further investigation.

The recessive character of the Cape Verdean mutation raises the prospect of identifying patients who carry two different (milder) *FUS* mutations. Such a possibility is especially intriguing, because it would reflect a mechanism common in complex genetic diseases that obscures the Mendelian patterns of inheritance and consequently complicates the detection of genetic defects.

It could also be that mechanistically different *FUS* mutations occur in related neurodegenerative brain disorders. Frontotemporal lobar degeneration, for example, is a neurodegenerative dementia that arises relatively frequently in patients with ALS, shares with ALS the characteristic TDP-43 inclusions in motor neurons, and co-occurs in families with ALS more often than expected by chance<sup>9</sup>. It will be interesting to see if *FUS* inclusions can be

identified in patients with frontotemporal lobar degeneration, with or without ALS. Although the carriers of the *FUS* mutations identified so far show no symptoms of the cognitive dysfunction characteristic of frontotemporal lobar degeneration, Kwiatkowski and colleagues' tissue autopsy analysis<sup>1</sup> revealed cytoplasmic *FUS* inclusions in neurons of the frontal cortex — the brain region affected in frontotemporal lobar degeneration.

With both *FUS* and TDP-43 mutations implicated in ALS, a pattern emerges that links abnormal RNA metabolism with motor-neuron disease. Although the latest findings<sup>1,2</sup> are based on an analysis of relatively few patients, they will have a much broader impact, not only by opening up new avenues of research, but also by accelerating efforts to prevent or ameliorate this devastating disease. ■

Kristel Slegers and Christine Van Broeckhoven are in the Neurodegenerative Brain Diseases Group, Department of Molecular Genetics, VIB, and in the Laboratory of Neurogenetics, Institute Born-Bunge, University of Antwerp, Antwerp B-2610, Belgium.

e-mail: christine.vanbroeckhoven@molgen.vib-ua.be

1. Kwiatkowski, T. J. Jr *et al.* *Science* **323**, 1205–1208 (2009).
2. Vance, C. *et al.* *Science* **323**, 1208–1211 (2009).
3. Valdmann, P. N. & Rouleau, G. A. *Neurology* **70**, 144–152 (2008).
4. Rosen, D. R. *et al.* *Nature* **362**, 59–62 (1993).
5. Sreedharan, J. *et al.* *Science* **319**, 1668–1672 (2008).
6. Sapp, P. C. *et al.* *Am. J. Hum. Genet.* **73**, 397–403 (2003).
7. Ruddy, D. M. *et al.* *Am. J. Hum. Genet.* **73**, 390–396 (2003).
8. Law, W. J. *et al.* *Brief Funct. Genomics Proteomics* **5**, 8–14 (2006).
9. Neumann, M. *et al.* *Science* **314**, 130–133 (2006).
10. Cironi, L. *et al.* *PLoS ONE* **7**, e2634 (2008).
11. Simpson, C. L. *et al.* *Hum. Mol. Genet.* **18**, 472–481 (2009).

## GEOCHEMISTRY

# A glacial hangover

Louis A. Derry

**The marine geochemical budget of some solutes does not add up. A test case shows that at least part of the reason may lie in the timescale over which continental weathering recovers from glaciations.**

On page 493 of this issue, Vance *et al.*<sup>1</sup> describe how they have tackled one of the puzzles in geochemistry: the fact that current tallies of the sources and sinks of certain elements and isotopes in the ocean don't match the observed rates of change. Vance *et al.* consider the case of strontium isotopes, where the isotopic ratio of <sup>87</sup>Sr/<sup>86</sup>Sr has been evolving at a faster rate than it should, given the known sources and sinks in the oceans. Their conclusion is that the imbalance lies on the river-input side of the equation, and they propose that estimates based on modern river data do not adequately account for the recovery from recent ice ages.

The Quaternary era, which began about 1.8 million years ago, is characterized by cycles between glacial periods of extensive continental ice sheets and the warmer, interglacial states that follow ice-sheet collapse. Over the past million years or so, the frequency of this cycle has been 100,000 years, which implies that processes operating at Earth's surface with timescales near that period or longer cannot reach steady state. Many solutes in the ocean have response times that are longer than the ice-age rhythm, and establishing their budgets is complicated by the fact that an assumption of a modern steady state does not apply. Although we cannot measure the past concentrations of ocean solutes, the isotopic compositions of solute elements can be recorded in marine sediments. Many show variations in time that can provide insight into the shifting balance of inputs to the oceans that occur in

response to tectonic or climatic factors.

The isotopic composition of strontium has been of particular interest because its temporal variations primarily reflect the long-term balance between weathering on land, which supplies solutes with high <sup>87</sup>Sr/<sup>86</sup>Sr, and submarine hydrothermal activity, whose net effect

is to supply strontium with low <sup>87</sup>Sr/<sup>86</sup>Sr. It has been recognized for some time that the isotopic budget of strontium in the oceans is difficult to balance with the known sources from rivers and input from hydrothermal vents at mid-ocean ridges<sup>2</sup>. The best estimates of the modern river and mid-ocean-ridge hydrothermal fluxes predict a rate of change in <sup>87</sup>Sr/<sup>86</sup>Sr in sea water that is much faster than observed. Either the flux of <sup>87</sup>Sr from rivers has been higher than we think, or the supply of <sup>86</sup>Sr from hydrothermal exchange is underestimated, or perhaps some of both<sup>3</sup>.

The river data represent a kind of weighted average from many rivers around the world, and it is possible that sampling bias has left us with an incomplete picture of river fluxes. It is also possible that the estimates of modern river



**Figure 1 | Earth moving.** Glacial action produces huge amounts of crushed and ground rock material, seen here as strips of moraine, that Vance *et al.*<sup>1</sup> argue is subject to enhanced weathering in the early period of glacial retreat.

P. A. SOUDERS/CORBIS



fluxes are in fact reasonably good, but that we are living in a time that is not 'typical' of the fluxes averaged over a longer timescale. This is the explanation explored by Vance and colleagues<sup>1</sup>: they propose that much of the world's land surface is suffering from a glacial hangover, and that geochemical fluxes have not yet recovered from the major perturbation of the last glacial–interglacial transition.

The major glaciations created tremendous volumes of finely ground but chemically fresh sediment, known as loess, which wind and water have by now distributed widely across the planet. Some of the major breadbaskets of the world, including the mid-continent North American and western Russian grain belts, occur on mineral-rich loess soils. But with time, soils become increasingly weathered and nutrient-poor, as is the case for many lowland tropical regions. Freshly ground rock material weathers rapidly, but the weathering rate declines over time. So, as the great ice sheets retreated, they left vast tracts of the Northern Hemisphere covered in fresh, finely ground rock (Fig. 1) that would have initially weathered fast, and been an excellent source of solutes with high  $^{87}\text{Sr}/^{86}\text{Sr}$ . But within a few thousand years, weathering rates would have declined noticeably.

Vance and colleagues apply a power law to quantify the relationship between substrate age and weathering rate that seems to satisfy, at least in broad terms, observations from several different environments. They also note that, in recently glaciated environments, not only does the amount of strontium released during weathering of granitic rocks change with time, but so does the isotopic ratio of the strontium released. Initially, breakdown of the mineral biotite releases strontium with a particularly high  $^{87}\text{Sr}/^{86}\text{Sr}$  ratio, but this effect declines with time<sup>4</sup>. So, as landscapes recover from glaciation, they release less strontium, and that strontium contains proportionally less of the  $^{87}\text{Sr}$  isotope, and the landscapes thus become less potent drivers of the marine budget. The timescale of this 'relaxation' is of the order of 10,000 years: although these landscapes have partially recovered from the effects of the last glaciation, they would have been packed a notably bigger punch 5,000 or 10,000 years ago.

Using these time-dependent relationships between soil age, and the amount and isotopic composition of strontium released by weathering, Vance *et al.* then compute the expected impact of the 100,000-year glacial–interglacial cycle on the isotopic budget of strontium in the oceans. Because strontium has a long residence time compared with the climate cyclicity, the  $^{87}\text{Sr}/^{86}\text{Sr}$  ratio does not show significant observable variation at the 100,000-year driving frequency: the damping effect of the long response time is too strong. But the authors do find that they can resolve much of the apparent discrepancy between current best estimates of the river and hydrothermal fluxes and the long-term strontium isotopic budget of the

oceans by accounting for variation in both the weathering flux and its isotopic composition on a glacial–interglacial timescale.

The non-steady-state weathering 'glacial hangover' hypothesis has implications for the cycles of other elements — including those, such as osmium, that are used as tracers; major salts such as magnesium and calcium; and carbon dioxide. Recognition that the present-day river-flux measurements are inadequate to describe the time-integrated flux to the oceans both complicates and clarifies our understanding of important geochemical cycles and how they respond to perturbations.

However, the non-steady-state weathering proposed by Vance *et al.* could be only part of the story. A source of strontium that is not well quantified is the low-temperature hydrothermal alteration of the oceanic crust, which occurs away from the axes of mid-ocean ridges. This flux is more widespread than was once thought. But it is difficult to quantify because it is not associated with the obvious chemical plumes of the famous high-temperature (350°C) black smokers along the axes of mid-ocean ridges, where most element exchange between the ocean and Earth's crust is believed to take place. There is good evidence that this off-axis flux is important for strontium — and perhaps even enough to close the isotopic mass balance<sup>5</sup>.

It is likely that both the mechanism proposed by Vance *et al.* and low-temperature alteration of the oceanic crust have a role in creating the apparent imbalance for strontium, but it is too early to reach firm conclusions on the absolute importance of each. Further analyses that involve other tracers, such as calcium isotope ratios, and particularly those with response times short enough to be sensitive to glacial cycles, such as silicon and osmium isotope ratios, can provide insight into the problem. But it may be that only a substantial improvement in measurement technology will allow us to truly resolve the relative impact of each flux, by providing data on the oh-so-subtle variation of the  $^{87}\text{Sr}/^{86}\text{Sr}$  ratio at the glacial–interglacial timescale. ■

Louis A. Derry is in the Biogeochemistry and Environmental Biocomplexity Program, Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, New York 14853, USA. e-mail: lad9@cornell.edu

1. Vance, D., Teagle, D. A. H. & Foster, G. L. *Nature* **458**, 493–496 (2009).
2. Hodell, D. A., Mead, G. A. & Mueller, P. A. *Chem. Geol.* **80**, 291–307 (1990).
3. Galy, A., France-Lanord, C. & Derry, L. A. *Geochim. Cosmochim. Acta* **63**, 1905–1925 (1999).
4. Blum, J. D. & Erel, Y. *Geochim. Cosmochim. Acta* **61**, 3193–3204 (1997).
5. Butterfield, D. A., Nelson, B. K., Wheat, C. G., Mottl, M. J. & Roe, K. K. *Geochim. Cosmochim. Acta* **65**, 4141–4153 (2001).

## STRUCTURAL BIOLOGY

# Spliceosome subunit revealed

Charles C. Query

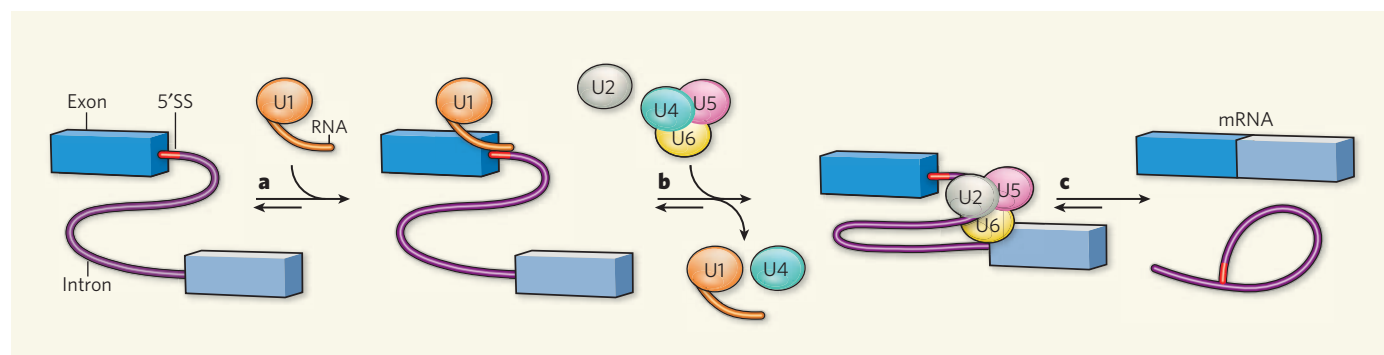
**The spliceosome enzyme binds to RNA transcripts at splice sites and removes intron sequences. The crystal structure of a spliceosome subunit shows how the enzyme recognizes one end of the intron.**

Among other defining features, such as the presence of a nuclear membrane separating genetic material from other parts of the cell, all eukaryotes examined to date contain at least one intron (a sequence that is spliced out from RNA transcripts). This is mirrored in all eukaryotic cells by the presence of parts of the spliceosome — the multi-component enzyme responsible for excising introns out of RNA and connecting together the remaining sequences (exons) to form a functional messenger RNA that can encode proteins. Almost all vertebrate genes contain introns, often in large numbers; variations in RNA splicing can thus generate a great diversity of protein amino-acid sequences.

The spliceosome is therefore implicated in protein evolution and is crucial to the storage and retrieval of genetic information in eukaryotes. It consists of five subunits known as small nuclear ribonucleoproteins (snRNPs), each one of which is a complex of a small RNA

and up to dozens of proteins. On page 475 of this issue, Nagai and colleagues<sup>1</sup> report the crystal structure of a spliceosomal snRNP called U1, whose core consists of an RNA and ten proteins. U1 snRNP is thought to form the first spliceosomal interaction at the 5' end of the intron and to stimulate the assembly of the rest of the spliceosome (Fig. 1). It was first implicated in splicing on the basis of sequence complementarity between one end of U1 RNA and the 5' boundary of introns<sup>2,3</sup>. The interactions between U1 components revealed by Nagai and colleagues' crystal structure not only suggest how the snRNP is assembled, but also hint at how the 5' splice site is recognized.

The authors<sup>1</sup> reconstituted U1 snRNP both from genetically modified versions of its constituent proteins (expressed in the bacterium *Escherichia coli*) that lacked various unstructured regions found in the normal proteins, and from RNA produced *in vitro*, thus obtaining crystals for their study. The Nagai group had



**Figure 1 | Spliceosome assembly and RNA splicing.** RNA transcripts consist of regions that encode proteins (exons) and non-coding regions (introns). Introns are excised by a multi-component ribonucleoprotein known as the spliceosome, which assembles at both ends of the intron. **a**, In the first step of assembly, the U1 component of the spliceosome binds to the 5' splice site (5'SS) through a combination of RNA–RNA and RNA–protein interactions. **b**, The remaining subunits of the

spliceosome (U2, and a complex of U4/5/6) assemble around the U1–intron complex, and then U1 and U4 are displaced, activating the catalytic spliceosome and RNA splicing. **c**, The spliceosome removes the intron as a lariat, and joins the exons together to form messenger RNA ready for translation. The crystal structure of U1 reported by Nagai and colleagues<sup>1</sup> offers clues to how U1 recognizes 5'SS and how U1 assembles from its own subunits.

previously solved the structures of several subunits of U1 snRNP<sup>4,5</sup>, and the present structure builds on the foundations of those studies.

The authors' findings<sup>1</sup> provide many insights into RNP structure and the interactions of RNPs with the 5' splice site (5'SS) of introns. The U1 RNA consists of four hairpin loops, plus an additional short helix; four of these five RNA duplexes form stacked helices that cross at approximately 90° to each other (see Fig. 1 on page 475), and the remaining hairpin structure is separated from these helices by a region known as the Sm binding site. This site is bound by the Sm (or LSm) complex, a common component of many snRNPs that is formed from seven proteins. The Nagai group had previously proposed<sup>4</sup> that the Sm complex is ring-shaped. The new structure<sup>1</sup> fully confirms this model, and additionally shows the contacts formed by the complex on U1 snRNA: the seven proteins each contact one nucleotide of RNA, which is held in a pseudo-helical arrangement. The interactions observed between the Sm complex and the RNA of U1 will almost certainly be typical of all the other Sm-containing snRNPs.

The fundamental role of U1 in 5'SS recognition has long been known, raising questions about how this recognition is mediated. Early studies<sup>6</sup> clearly showed base-pairing between the 5' end of U1 RNA and the 5'SS sequence; later evidence<sup>7</sup> revealed that the U1 protein U1C (which does not contribute to the Sm complex) also recognizes RNA sequences characteristic of 5'SS. It remained unclear how, or whether, both of these parts of U1 worked together to provide sequence-specific recognition.

Nagai and colleagues<sup>1</sup> now provide a structural rationale for the roles of both U1C and U1 RNA in 5'SS recognition. Although there was no authentic intron 5'SS present in the authors' crystal, the 5' end of each U1 RNA was partially base-paired to the 5' end of another U1 from an adjacent snRNP complex. This binding provides a good first-approximation mimic for U1–5'SS interaction. The authors observe

that certain amino-acid residues in the U1C protein interact with the 'minor groove' of the U1–5'SS duplex, approximately at the position of a dinucleotide motif that is found at the exon–intron boundary of most 5'SSs (Fig. 1).

Although the resolution of Nagai and colleagues' current structure is insufficient to observe details of the U1C–RNA interaction, it is likely that U1C helps to stabilize the duplex. This is consistent with previous observations<sup>8,9</sup> regarding an ATPase enzyme that is thought to disrupt U1C–5'SS interactions: mutations to U1C that weaken U1C–5'SS interactions compensate for the defective action of a mutant of the ATPase enzyme. It will be interesting to see in future structures whether there are additional RNA–protein contacts that help stabilize the U1–intron complex. For example, do any of the RNA sequences flanking the 5'SS touch U1? And do any of the unstructured regions that were omitted from the proteins incorporated into Nagai and colleagues' snRNP form additional contacts with the intron?

Another intriguing aspect of the authors' structure<sup>1</sup> is the long-range protein connectivity within the complex. One of the U1 proteins, known as U1-70K, forms an unexpectedly long helix that lies alongside one of the U1 RNA helices. Another part of U1-70K wraps approximately halfway around the rest of the snRNP complex, and extends to the U1C protein, thus connecting sites that are essentially at distant, opposite ends of U1 (see Fig. 5a on page 479). This connection with the U1C protein helps to explain why U1C recruitment to the snRNP depends on the presence of U1-70K.

As always, there is much to look forward to in the future. For example, the binding of U1 to 5'SS stimulates other spliceosome-assembly events and other RNA-processing reactions — but how? Does U1 undergo conformational changes on binding to an intron that allow interactions with other snRNPs and complexes? The long helices and unstructured strands threading throughout the U1 structure are reminiscent, for example, of the long

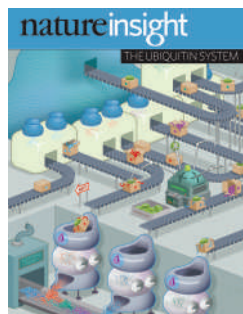
helices found in MutS (refs 10,11), a protein that repairs defects in DNA caused by mismatched bases. These helices are thought to help transmit a conformational change from one part of the protein to another when MutS binds to ATP molecules (the main source of energy in cells). Future structures of U1 with and without a bound 5'SS (or with and without a 5' end of U1 RNA) might help to predict the changes in the global U1 structure that occur on 5'SS binding, and how these changes affect the ability of U1 to interact with other spliceosomal components.

It is striking that, because U1 snRNP is not a compact, globular structure, it might be able to form multiple simultaneous contacts with other biological machinery. U1 snRNP (or its components) is known to associate with several other RNA-processing complexes<sup>12–14</sup>. Future studies should help to reveal features of U1 that participate in these additional interactions. In the meantime, the present structure<sup>1</sup> of U1 snRNP provides a rich array of information that will stimulate studies of 5'SS recognition, snRNP biogenesis, snRNP–snRNP interactions and the coordination of nuclear events. ■

Charles C. Query is in the Department of Cell Biology, Albert Einstein College of Medicine, Bronx, New York 10461-1975, USA.  
e-mail: query@einstein.yu.edu

1. Pomeranz Krummel, D. A., Oubridge, C., Leung, A. K. W., Li, J. & Nagai, K. *Nature* **458**, 475–480 (2009).
2. Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L. & Steitz, J. A. *Nature* **283**, 220–224 (1980).
3. Rogers, J. & Wall, R. *Proc. Natl Acad. Sci. USA* **77**, 1877–1879 (1980).
4. Kambach, C. et al. *Cell* **96**, 375–387 (1999).
5. Oubridge, C., Ito, N., Evans, P. R., Teo, C. H. & Nagai, K. *Nature* **372**, 432–438 (1994).
6. Zhuang, Y. & Weiner, A. M. *Cell* **46**, 827–835 (1986).
7. Du, H. & Rosbash, M. *Nature* **419**, 86–90 (2002).
8. Chen, J. Y. et al. *Mol. Cell* **7**, 227–232 (2001).
9. Staley, J. P. & Guthrie, C. *Mol. Cell* **3**, 55–64 (1999).
10. Lamers, M. H. et al. *Nature* **407**, 711–717 (2000).
11. Obmolova, G., Ban, C., Hsieh, P. & Yang, W. *Nature* **407**, 703–710 (2000).
12. Kwek, K. Y. et al. *Nature Struct. Biol.* **9**, 800–805 (2002).
13. Lewis, J. D., Izaurralde, E., Jarmolowski, A., McGuigan, C. & Mattaj, J. W. *Genes Dev.* **10**, 1683–1698 (1996).
14. Shi, Y. et al. *Mol. Cell* **33**, 365–376 (2009).



**Cover illustration**

In the ubiquitin system, proteins are either targeted to the proteasome for degradation or sent to different locations in the cell. (Artwork by N. Spencer.)

**Editor, *Nature***

Philip Campbell

**Publishing**

Nick Campbell  
Claudia Banks

**Insights Editor**

Lesley Anson

**Production Editor**

Davina Dadley-Moore

**Senior Art Editor**

Martin Harrison

**Art Editor**

Nik Spencer

**Sponsorship**

Amélie Pequignot  
Reya Silao

**Production**

Jocelyn Hilton

**Marketing**

Elena Woodstock  
Emily Elkins

**Editorial Assistant**

Emma Gibson

# THE UBIQUITIN SYSTEM

**T**he destruction of proteins is as important as their synthesis for the maintenance of protein homeostasis in cells. In eukaryotes, the ubiquitin–proteasome system is responsible for most of this protein degradation: the small protein ubiquitin acts as a death warrant, tagging and targeting other proteins to the large proteolytic chamber of the proteasome.

The discovery in the 1970s that certain proteins are ubiquitinated before degradation was awarded the 2004 Nobel Prize in Chemistry. It is now known that ubiquitin-mediated destruction plays a crucial part in cell-cycle regulation, DNA repair, cell growth and immune function, as well as in hormone-mediated signalling in plants. More recently, ubiquitin has been shown to have numerous non-proteolytic functions, including involvement in vesicular trafficking pathways, regulation of histone modification and viral budding.

Given the central role of the ubiquitin system in diverse cellular processes, it is not surprising that its dysfunction contributes to cancer and to neurodegenerative and immunological disorders. An understanding of the ubiquitin system is therefore important in devising treatments for such diseases.

With topics as diverse as the origin of the ubiquitin system and cancer therapy targeting the ubiquitin pathway, this Insight provides both an introduction and an update to the most topical themes in ubiquitin research.

We are pleased to acknowledge the financial support of Hybrigenics, ITI Life Sciences and Scil Proteins in producing this Insight. As always, *Nature* retains sole responsibility for editorial content and peer review.

Deepa Nath, Senior Editor

Sadaf Shadan, Associate Editor, News & Views

## REVIEWS

### 422 Origin and function of ubiquitin-like proteins

M. Hochstrasser

### 430 Ubiquitylation in innate and adaptive immunity

V. G. Bhoj & Z. J. Chen

### 438 Targeting the ubiquitin system in cancer therapy

D. Hoeller & I. Dikic

### 445 The ESCRT machinery in endosomal sorting of ubiquitylated membrane proteins

C. Raiborg & H. Stenmark

### 453 The ubiquitylation machinery of the endoplasmic reticulum

C. Hirsch, R. Gauss, S. C. Horn, O. Neuber & T. Sommer

### 461 Principles of ubiquitin and SUMO modifications in DNA repair

S. Bergink & S. Jentsch

nature  
insight

# Origin and function of ubiquitin-like proteins

Mark Hochstrasser<sup>1</sup>

**Eukaryotic proteins can be modified through attachment to various small molecules and proteins. One such modification is conjugation to ubiquitin and ubiquitin-like proteins (UBLs), which controls an enormous range of physiological processes. Bound UBLs mainly regulate the interactions of proteins with other macromolecules, for example binding to the proteasome or recruitment to chromatin. The various UBL systems use related enzymes to attach specific UBLs to proteins (or other molecules), and most of these attachments are transient. There is increasing evidence suggesting that such UBL-protein modification evolved from prokaryotic sulphurtransferase systems or related enzymes. Moreover, proteins similar to UBL-conjugating enzymes and UBL-deconjugating enzymes seem to have already been widespread at the time of the last common ancestor of eukaryotes, suggesting that UBL-protein conjugation did not first evolve in eukaryotes.**

The proteins in a eukaryotic cell are subject to a huge variety of post-translational modifications, which greatly extend the functional diversity and dynamics of the proteome. Proteins can be modified by attaching to small molecules such as phosphate groups, methyl groups or acetyl groups, or, usually only transiently, to certain proteins<sup>1–3</sup>. The first such protein-based modification to be described was ubiquitin, and this is now the most thoroughly understood. Ubiquitin is a small protein that is extremely well conserved among the Eukaryota but absent from members of the other two superkingdoms, the Eubacteria and the Archaea. It can be transiently attached to thousands of different proteins.

An intricate enzymatic system catalyses the modification of substrate proteins with ubiquitin, a process known as ubiquitylation (Box 1). In a similar manner, distinct but evolutionarily related enzyme cascades catalyse the attachment of ubiquitin-like proteins (UBLs) to proteins or other molecules<sup>4</sup>. As ubiquitin and UBLs have the same three-dimensional core structure — the  $\beta$ -grasp fold — it is clear that the different UBL-modification systems have a common ancestry.

For the first few decades after the discovery of ubiquitin in 1975, the evolutionary precursors to the ubiquitin-conjugation system went largely undetected. Ubiquitin itself is regarded as one of the most highly conserved of all eukaryotic proteins<sup>5</sup>, but until relatively recently sequence-comparison algorithms were not sensitive enough to detect any bacterial proteins with amino-acid sequence similarity. This situation has changed considerably in the past few years. First, sequence-comparison methods have become more sophisticated, uncovering unexpected similarities between ubiquitin-pathway components (including ubiquitin) and various bacterial proteins<sup>6</sup>. Second, structure determinations have shown that a ubiquitin-like fold is adopted by many eukaryotic and prokaryotic proteins (or their domains), even when amino-acid sequence similarity to ubiquitin is minimal. Third, mechanistic analysis of the biosynthesis of secondary metabolites and enzyme cofactors such as thiamine (vitamin B<sub>1</sub>) in prokaryotes has uncovered parallels to the activation and conjugation of UBLs<sup>7</sup>. Together, these studies imply that the ubiquitin system has been ‘cobbed together’ from a variety of pre-existing parts and pathways that had already undergone considerable diversification in prokaryotes<sup>1,6</sup>.

The diversity of the processes regulated by ubiquitin-protein modification is extraordinary<sup>4,8</sup> and the consequences of the

modification depend on whether the ubiquitin is attached to the protein as a monomer or as polyubiquitin chains (Fig. 1). Within the chains, the different ubiquitin-ubiquitin linkages help to dictate the fate of the modified substrate. Often when a substrate protein is coupled to a polyubiquitin chain, it binds to the 26S proteasome, a large multisubunit protease complex that degrades the substrate into small peptides and recycles the ubiquitin tag. This happens for substrates that must be eliminated for proper cell-cycle progression, transcriptional regulation, protein quality control, signal transduction or circadian rhythms. Ubiquitylation is also used in non-proteolytic regulatory mechanisms, such as membrane-protein endocytosis and intracellular trafficking, chromatin-mediated regulation of transcription, DNA repair and the assembly of signalling complexes. In this light, it is not surprising that the list of diseases implicating misregulation of the ubiquitin system is growing steadily, and currently includes many cancers, some severe types of mental retardation (such as Angelman syndrome), neurodegenerative disorders such as Parkinson's disease, Huntington's disease and Alzheimer's disease, and type 2 diabetes<sup>9</sup>.

In this Review, I discuss the basic features of UBL conjugation to proteins and how such modifications contribute to cellular regulatory mechanisms. I focus on the probable evolutionary antecedents to eukaryotic UBL-protein conjugation in prokaryotes but do not discuss the evolution of the proteasome as it has been well covered elsewhere<sup>10,11</sup>. I also look at the parallels between enzymes of the ubiquitin system and those that are responsible for the synthesis of specific small molecules (particularly certain sulphurtransferases). In light of the recent wave of research on these evolutionary and mechanistic questions, it seems appropriate to review some of the advances made.

## UBLs and related protein domains

Studies begun in the late 1980s identified an interferon-stimulated gene product of 15 kDa (ISG15) that has significant sequence similarity to ubiquitin and also covalently modifies other proteins<sup>12,13</sup>. ISG15 turned out to be the first in a number of UBLs found to function as protein modifiers (Table 1). Despite being the first UBL to be identified, the function of ISG15 is still poorly understood, and its E1-like (ISG15-activating) and E2-like (ISG15-conjugating) enzymes (Box 1) were not identified until much later<sup>14–17</sup>. These enzymes, like ISG15 itself, are strongly induced by type I interferons. Results from studies of mouse models have indicated that the conjugation of ISG15 to proteins

<sup>1</sup>Yale University, Department of Molecular Biophysics & Biochemistry, 266 Whitney Avenue, PO Box 208114, New Haven, Connecticut 06520, USA.



contributes to antiviral responses, consistent with ISG15 being induced by type I interferons<sup>18–20</sup>, which are some of the first factors produced by the innate immune system in response to viruses.

Like ubiquitin, nine UBLs have been shown to covalently modify other macromolecules, usually proteins (Table 1), and several other factors are thought to have this ability. This list of UBLs is probably incomplete. Ubiquitin modifies more than 1,000 different proteins in yeast<sup>21</sup>. Some UBLs, such as SUMO (small ubiquitin-related modifier), might target a similar number and diversity of substrates, but others have a much more limited range of substrates than does ubiquitin<sup>4</sup>. For example, the yeast protein Atg12 seems to have a single target (Atg5), and Atg8 attaches to a specific phospholipid, phosphatidylethanolamine<sup>22</sup>.

As noted earlier, most UBL-modification pathways use similar enzymatic mechanisms. The main pathways of protein conjugation seem to have evolved through repeated rounds of duplication and diversification of enzymes and protein modifiers derived from ancient biosynthetic pathways (see below). However, a few unusual UBL-conjugation mechanisms have been proposed for specific UBL pathways. In one of these models, a ubiquitin hydrolase was proposed to be able not only to cleave ubiquitin from substrates but also to work backwards, in effect, and ligate ubiquitin to a protein<sup>23</sup>. Another unconventional model came from the results of sequence analysis of an unusual group of putative self-splicing polyproteins in ciliates<sup>24</sup>. These polyproteins consist of a tandem series of variant UBL domains with interspersed self-splicing bacterial intein-like (BIL) domains; the corresponding genes probably arose from a polyubiquitin-encoding gene that had acquired a sequence encoding the BIL domain.

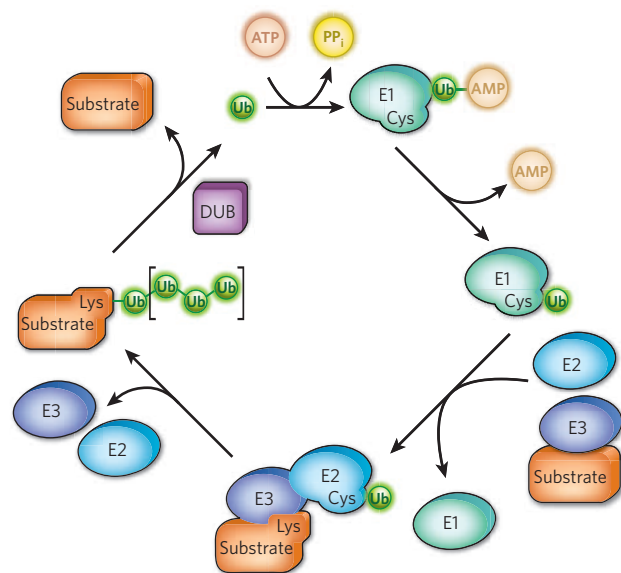
BIL domains have a serine or cysteine residue at their amino terminus, and they activate a peptide-bond rearrangement in which the upstream peptide acyl chain is transferred from N to O (for serine) or from N to

S (for cysteine) on the terminal BIL amino acid. This rearrangement initiates cleavage and self-splicing reactions<sup>24</sup>. The acyl shifts (N→S shifts) triggered by the BIL domains in the BIL–ubiquitin-like (BUBL) proteins might facilitate nucleophilic attack on the resultant thioester by a protein (or other molecule) during autocatalytic processing, leading to ligation of the upstream UBL sequence to the attacking molecule<sup>24</sup>. If the attacking group was a lysine side chain of a protein, the resultant product would be a UBL–protein conjugate similar to those formed through the standard pathway (Box 1), even though no E1 or E2 (or ATP) is involved. Notably, the sequences upstream of the BIL domains in BUBL precursors do not need to be UBLs. Thus, there could be a class of protein modifiers that are not related to ubiquitin but are fused upstream of domains with a similar function to BIL domains.

There are also many ubiquitin-related proteins in which the ubiquitin-like domain (ULD) is part of a larger polypeptide but, usually, is neither processed nor covalently attached to other proteins<sup>1</sup>. Such ULDs confer properties on a protein that are similar to those from a transferable UBL, including the ability to bind to specific target proteins. Some ULDs can be cleaved under specific conditions (and some might even become competent for ligation to other proteins). For example, autocleavage at an internal ULD occurs in the deubiquitylating enzyme USP1 after damage to cells by ultraviolet light. The damage thus leads to inactivation of the enzyme and allows the accumulation of monoubiquitylated proliferating cell nuclear antigen (PCNA), which is required for *trans*-lesion DNA synthesis<sup>25</sup>.

The β-grasp fold shared by all structurally characterized UBLs and ULDs is phylogenetically widespread and ancient; it may have arisen as an RNA-binding module in a primitive protein-translation system<sup>26</sup>. It has been adapted to a broad array of functions, ranging from a scaffold for various enzymatic activities and iron–sulphur-cluster binding to an adaptor module for specific protein–protein interactions.

#### Box 1 | The basics of the ubiquitin-conjugation pathway



Ubiquitin (green circles) is a 76-residue polypeptide that is activated and attached to substrate proteins by a series of enzymes (see box figure). The modification of proteins with ubiquitin-like proteins (UBLs) also follows the same general mechanisms.

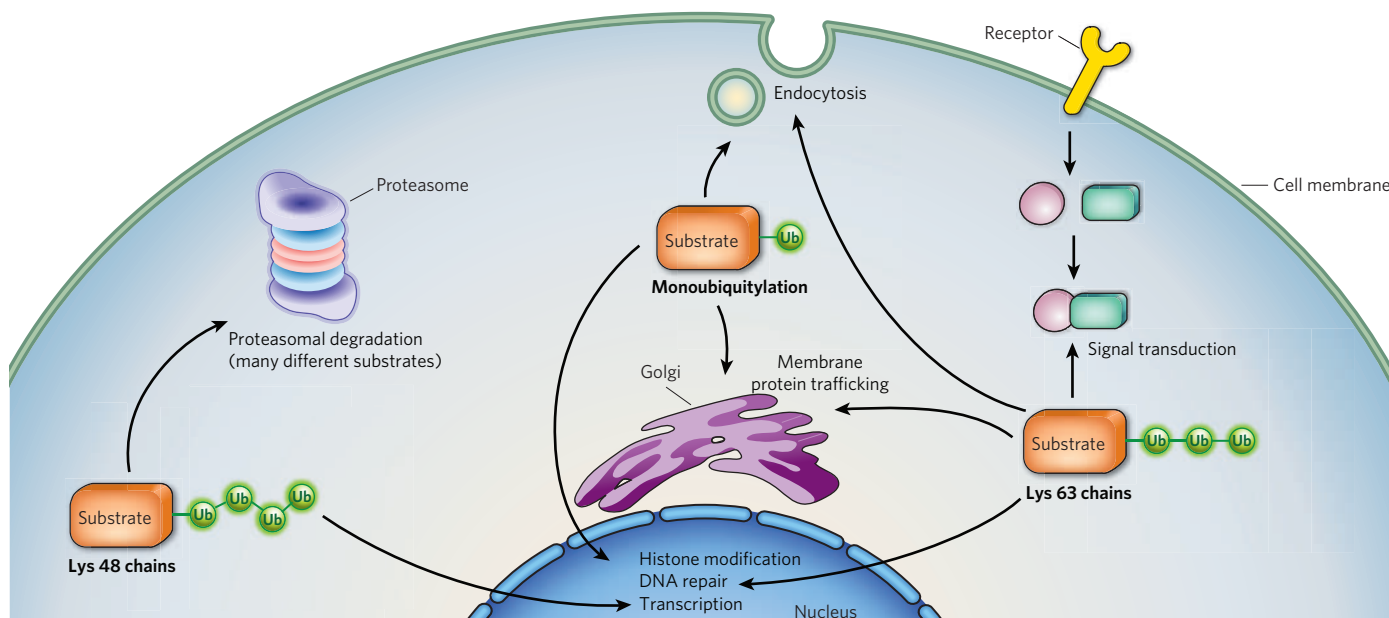
Three types of enzyme — E1, E2 and E3 — carry out ubiquitin-modification reactions, including the assembly of polyubiquitin chains (conjugation of additional ubiquitins to a single ubiquitin to form a polyubiquitin chain on some proteins is indicated by the brackets in the image). E1s activate ubiquitin. E2s pick up the ubiquitin by transthioesterification from E1 and conjugate it to substrates. E3s then ligate the ubiquitin to the substrate (and in some cases, form an intermediary thioester with ubiquitin before the final transfer to the substrate). All eukaryotes

encode multiple isozymes of E2 and E3, up to several dozen E2s and many hundreds of E3s. This allows the modification of many proteins in a highly specific manner, and such modifications are often under strict temporal and spatial control.

Ubiquitin is usually attached to proteins through an amide linkage between its carboxyl terminus and a primary amino group on the acceptor protein<sup>2,52</sup>. The most common linkages are with the ε-amino group (side chain) of a lysine residue, but linkages can also form with the amino group at the amino terminus of the substrate<sup>70</sup>. In addition, ubiquitin has recently been found to be attached to cysteine, serine and threonine residues in proteins<sup>71–73</sup>. In ubiquitin polymers, the lysine side chain of one ubiquitin molecule is linked to the C terminus of another ubiquitin molecule, and so on. Ubiquitin contains seven lysine residues, all of which can contribute to such linkages.

Ubiquitin has a glycine residue at its C terminus, and this must be activated before it can form a covalent bond with another protein<sup>2,52</sup>. Initially, the C terminus is adenylated by E1, with the ubiquitin–AMP adduct remaining bound to the enzyme. A cysteine side chain in E1 then attacks the C terminus of ubiquitin, yielding an E1–ubiquitin thioester intermediate. The activated ubiquitin is subsequently passed to a cysteine residue in the active site of an E2 (ref. 74). Together with an E3, the E2 catalyses ubiquitylation of the substrate. E3s have a crucial role in recognition of the substrate (although not all UBL pathways require an E3). In the ubiquitin pathway, a different E3 can help to attach ubiquitin molecules to a protein that has already been modified by one or more ubiquitins. Such E3s are sometimes called E4s, particularly when they are thought to extend a polyubiquitin chain. Enzymes known as deubiquitylating enzymes (DUBs) can remove ubiquitin molecules that are attached to proteins. As a result of DUB activities, ubiquitin modifies proteins only transiently<sup>61,75</sup>.

This process of dynamically modifying proteins with ubiquitin and other UBLs creates reversible switches between different functional states of a substrate protein, allowing fine control of numerous cellular pathways.



**Figure 1 | Cellular processes that depend on ubiquitin conjugation.** Protein attachment to a single ubiquitin (monoubiquitylation) allows recognition by a subset of ubiquitin-binding domains (UBDs) in target proteins, and this is important in the indicated cellular processes. During ubiquitylation, often a single, specific lysine is modified. Different polyubiquitin chains are thought to have different structures that allow discrimination by UBDs and deubiquitylating enzymes, although other contextual cues, such as the cellular location where the modification occurs, may also help to dictate

the physiological consequences of the polyubiquitin attachment. Lys-48-linked chains are most commonly associated with proteasomal binding and degradation, but they can also regulate transcription by a non-proteolytic mechanism. Lys-63-linked chains participate in signal transduction, membrane-protein trafficking, endocytosis and DNA repair. Substrate-specific monoubiquitylation regulates all of these processes as well as chromatin-modulated gene transcription. Not shown here are ubiquitin chains with mixed linkages or multisite ubiquitylation of the substrate.

### Consequences of UBL-protein conjugation

Early work on ubiquitin focused mainly on its role in proteolysis<sup>27–29</sup>. The 26S proteasome is responsible for the degradation of polyubiquitylated proteins, and direct binding of a polyubiquitin chain to the proteasome can fully account for the observed affinity of model polyubiquitylated proteins for this protease complex<sup>30</sup>. In short, the polyubiquitin chain provides a generic affinity tag that results in tight binding of the proteolytic substrate to the proteasome (Fig. 2). Multiple polyubiquitin receptors are present within the proteasome<sup>31–33</sup>. In addition, polyubiquitin-binding domains are found in mobile shuttling factors that direct polyubiquitylated proteins to the proteasome<sup>34–36</sup>.

Specific interactions between ubiquitin and ubiquitin-binding domains (UBDs) are not limited to the binding of ubiquitylated molecules to the proteasome. A general theme that has emerged over the past decade is that many of the functions of ubiquitin and UBLs are mediated by association with UBDs<sup>37</sup>. For instance, a UBD in the Vps23/TSG101 membrane-protein sorting factor mediates recognition of monoubiquitylated membrane proteins, and several distinct UBDs in the deubiquitylating enzyme Ubp14/USP5 allow it to specifically recognize unanchored polyubiquitin chains. At least 16 structurally distinct UBD classes have been characterized so far, and the domains in these classes vary considerably in size (from ~30 to 150 residues) and tertiary structure<sup>37</sup>. The evolution of these different types of UBD allowed a huge expansion of cellular processes that depend on ubiquitin conjugation in modern eukaryotes.

UBDs usually bind to ubiquitin only weakly<sup>37</sup>. Despite the generally low affinities, mutational studies support the physiological relevance of these associations. Linking multiple ubiquitin moieties into a chain can have marked effects on the affinity or avidity of ubiquitin for target proteins. A good example comes from an analysis of polyubiquitin–proteasome binding<sup>30</sup>. Cecile Pickart and colleagues<sup>30</sup> measured the extent to which the degradation of model substrates was inhibited in the presence of ubiquitin chains of various lengths. Tetrameric chains were strongly inhibitory. By contrast, inhibition was extremely weak in the presence of trimeric chains and undetectable with dimeric chains. The lack of a simple dependence of affinity on the length of the chain implies that

formation of a tetrameric chain creates a unique binding determinant that is not present in shorter chains or in monoubiquitin<sup>38</sup>.

Of the UBLs, SUMO is the most extensively studied. Analyses of the interactions between SUMO and other proteins confirm and extend many of the ideas about ubiquitin–protein interactions detailed so far, and findings from genetic, biochemical and biophysical studies have converged on a single, non-covalent SUMO-binding element, called the SUMO-interaction motif (SIM)<sup>39</sup>. SIM-bearing peptides as short as nine residues (much shorter than the domains that typically bind to ubiquitin) have dissociation constants of between 5  $\mu$ M and 10  $\mu$ M, which is substantially tighter binding than that seen with most known UBDs. The SIM consensus sequence contains a central group of three or four hydrophobic residues, usually with a cluster of acidic residues on one side. The SIM forms a  $\beta$ -strand that sits in a hydrophobic surface depression between the  $\beta_2$ -strand and  $\alpha_1$ -helix of SUMO; acidic residues at the ends of the SIM interact with basic residues on the SUMO surface<sup>40,41</sup>. This is the opposite face of the  $\beta$ -grasp fold from that recognized on ubiquitin by most UBDs; the surface recognized is a hydrophobic patch centred on the isoleucine residue at position 44 of ubiquitin. Thus, different UBLs can function analogously as adaptor modules, but the way in which they bind to target proteins need not be the same.

It is clear, therefore, that the conjugation of UBLs to a target protein frequently functions to promote interaction of the target with other proteins. When such modifications enhance a target's interaction with another macromolecule, they usually do so by participating directly in the formation of the binding interface with the target molecule (Fig. 3a). An interaction can also be modulated by an allosteric change in a target-binding site induced by the attached UBL (Fig. 3b). Most known cases of regulation by UBLs fall into the former category, and only a handful operate by the latter<sup>42,43</sup>. Even if only a small proportion of the cellular protein was modified with a UBL, the altered activity of the protein could be enough to effect a change in physiological state. Such non-covalent UBL–protein interactions tend to be weak, however (as mentioned earlier). Specific binding can be greatly enhanced either by polymerization of the ubiquitin or UBL signal, or by combining the weak binding to the



ubiquitin/UBL with additional weak binding sites in the conjugated protein (Fig. 3a, c). An example of such multivalent binding is the association of SUMO-conjugated RanGAP1 with the nuclear-pore complex; neither SUMO nor RanGAP1 alone can bind tightly to this complex<sup>44</sup>.

Given their bulk, another way by which UBLs can exert their function is through steric hindrance: the attached UBL can simply block the binding of one protein to another (or to another part of the same protein) (Fig. 3d). There are relatively few well-established *in vivo* examples of this intermolecular inhibitory mechanism. One likely reason is that for such a mechanism to operate effectively, a large proportion of the cellular protein would need to be modified by the UBL. However, for many proteins, only a minuscule amount is observed in conjugate form. In principle, such an inhibitory mechanism could still operate if the small proportion of modified protein were localized to a functionally privileged cellular site or if a transient modification were enough to shift the protein to a new state.

### Origins of UBL-protein conjugation

An early clue to the antecedents of UBL-protein conjugation came from the amino-acid sequence of the yeast ubiquitin-activating enzyme E1. E1 was found to have weak, but significant, sequence similarity to MoeB, an *Escherichia coli* protein that is required for biosynthesis of the molybdenum cofactor (Moco)<sup>45</sup>. But the biochemical function of MoeB was unknown when E1 was sequenced in 1991, so this similarity was not particularly informative. During the late 1990s, however, the protein sequences and catalytic mechanisms of the enzymes used to synthesize Moco (and thiamine) began to be deciphered, and intriguing similarities to ubiquitin activation were noted<sup>46–48</sup>. Because the thiamine synthesis pathway is present in nearly all bacterial species and Moco synthesis enzymes are also found in a broad range of bacteria, these enzymatic systems are thought to be evolutionarily much more ancient than ubiquitin conjugation<sup>6</sup>.

### UBL-related sulphur-carrier proteins

To synthesize the cofactors Moco and thiamine, sulphur atoms need to be inserted into their precursors. A small sulphur-carrier protein (MoaD in Moco synthesis and ThiS in thiamine synthesis) donates the

necessary sulphur(s). Sulphur is taken from a thiocarboxylate group formed at the C terminus of these proteins (Fig. 4). MoaD and ThiS are related and, like ubiquitin, their sequences end with a pair of glycine residues. Conversion of the C-terminal glycine carboxylate to a thiocarboxylate in these proteins is preceded by adenylation of the C terminus by an E1-related enzyme: MoeB for MoaD and ThiF for ThiS<sup>47–49</sup>. Like ubiquitin, both MoaD and ThiS have a  $\beta$ -grasp fold, despite having minimal sequence similarity to ubiquitin<sup>50,51</sup>. Therefore, ubiquitin, MoaD and ThiS are structurally related proteins with C termini that are activated through adenylation by homologous E1-like enzymes<sup>52,53</sup>.

In the early 2000s, further insight into the potential evolutionary link between these sulphur-transfer systems and UBL activation came from the discovery of ubiquitin-related modifier 1 (Urm1) (ref. 54), a protein first found in *Saccharomyces cerevisiae* that is related to MoaD and ThiS (Fig. 4, right). Although *S. cerevisiae* lacks any Moco-containing enzymes and uses a different mechanism to synthesize thiamine, it does express an E1-related protein, Uba4, which has sequence similarity to ThiF and MoeB. Uba4 binds to Urm1 and stimulates the covalent addition of Urm1 to cellular proteins<sup>54,55</sup>. These findings indicate that Urm1 and Uba4 function as part of a UBL-protein conjugation system, despite having much closer sequence relatedness to bacterial sulphur-transfer enzymes. However, as discussed below, Uba4 also functions in sulphur transfer, so it could be a bifunctional enzyme<sup>56</sup>. The Urm1–Uba4 system may therefore be a ‘molecular fossil’ that retains features of the more ancient sulphur-transfer pathway yet can also conjugate UBLs to proteins.

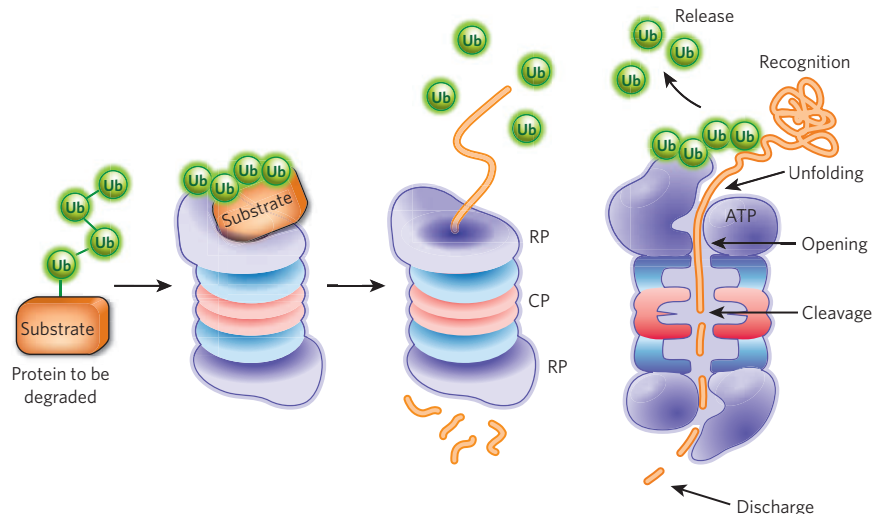
### Potential dual function of Urm1

The enzymatic activation of the C termini of UBLs by ATP-dependent E1-related enzymes is analogous to other adenylation reactions such as the activation of amino acids by aminoacyl-tRNA synthetases<sup>2</sup>. For UBLs, the energy of ATP hydrolysis is conserved through the formation of the E1–UBL thioester linkage. However, despite being chemically activated, all of the known UBLs (Table 1) except for Urm1 also undergo an energetically neutral transesterification to the thiol of a second enzyme, E2 (Box 1). The one apparent exception to the use of an E1–E2 relay, the Urm1 pathway, has distinct features, which it shares with bacterial sulphur-transfer pathways. Uba4 (the E1 in the Urm1

**Table 1 | Known and putative UBLs and their activating and conjugating enzymes**

UBL*	Identity with ubiquitin (%)	E1 (UBL-activating enzyme)*	E2 (UBL-conjugating enzyme)*	Comments on UBL
<b>Known UBLs</b>				
Ubiquitin	100	Uba1 (UBA6)	Many	Precursors encoded by multiple genes
Rub1 (NEDD8)	55	Uba3–Uba1 heterodimer	Ubc12	Substrates are cullins and p53
FUBI (also known as MNSF- $\beta$ or FAU)	38	NI	NI	Derived from a ribosomal-protein precursor
FAT10	32 and 40†	UBA6	NI	Contains a $\beta$ -grasp fold Substrates unknown
ISG15	32 and 37†	UBE1L	UBCH8	Production induced by type I interferons
Smt3 (SUMO1, SUMO2, SUMO3)	18	Uba2–Aos1 heterodimer	Ubc9	SUMO encoded by 3–4 genes in vertebrates, depending on the species
Atg8	ND	Atg7	Atg3	Three known isoforms in humans Contains a $\beta$ -grasp fold
Atg12	ND	Atg7	Atg10	~20% identical to Atg8
Urm1	ND	Uba4	NI	Related to the small sulphur-carrying proteins MoaD and ThiS Contains a $\beta$ -grasp fold
UFM1	ND	UBA5	UFC1	Contains a $\beta$ -grasp fold
<b>Putative UBLs</b>				
BUBL1, BUBL2	Variable (up to 80%)	NI	NI	Putative autoproteolytic proteins in ciliates
UBL-1	40	NI	NI	A precursor to ribosomal proteins in nematodes
SF3A120	30	NI	NI	UBL domain at C terminus No data about conjugation
Oligoadenylate synthetase	30 and 42†	NI	NI	UBL domain at C terminus No data about conjugation

ND, not detectable by standard BLAST searches. NI, not identified. \*UBLs are listed as the yeast (*Saccharomyces cerevisiae*) symbol if the UBL is present in yeast, otherwise vertebrate symbols are listed. Known vertebrate orthologues with symbols that differ from yeast proteins are listed in parentheses. For E1s and E2s, yeast symbols are listed if the protein is found in yeast. (UBA6 has a much more limited phylogenetic distribution than Uba1.) †The identities listed are for each of two ubiquitin-related domains.



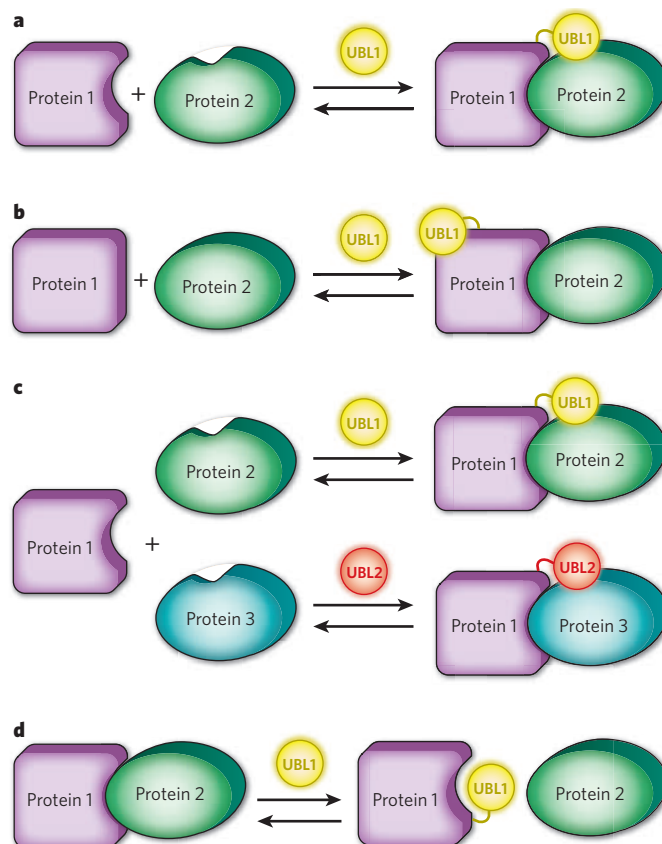
**Figure 2 | Polyubiquitin-tagged proteins are often targeted for proteasome-mediated degradation.** The ubiquitin-proteasome pathway is responsible for the degradation of hundreds, and probably thousands, of proteins. Many of these substrates are regulatory proteins, such as transcription factors or cell-cycle regulators; others are misfolded or otherwise aberrant proteins that must be eliminated to prevent their aggregation or toxicity. A polyubiquitin-modified protein is the form most commonly targeted to the proteasome. Ubiquitin receptors either in the proteasome regulatory particle (RP, purple) of the 26S proteasome or adaptor proteins that associate reversibly with both polyubiquitylated proteins and specific proteasomal subunits (not shown) allow binding of the proteolytic substrate to the proteasome. As shown in the cut-away on the right, ATPases within the RP unfold the substrate and translocate it into the 20S proteasome core particle (CP, blue and red rings), which houses the proteolytic sites in an interior chamber. The substrate is cleaved to small peptides. Ubiquitin itself is normally recycled by DUBs that bind to or are intrinsic to the RP.

pathway) contains a rhodanese-homology domain (RHD), unlike the E1s for other UBLs<sup>1,56</sup> (Fig. 4). Rhodanese and several RHD-containing proteins are sulphurtransferases, and they operate by transferring sulphur to targets by way of an intermediate persulphide (S-S-H) on their active-site cysteine residue<sup>57</sup>. Many MoeB-family proteins have a domain organization similar to that of Uba4, with an E1-like domain and, C terminal to this, an RHD. On the basis of these and other considerations, it had been proposed that thiocarboxylate formation in MoaD proceeds through an acyl disulphide intermediate between MoaD and the RHD of the MoeB-related enzyme, and that, by analogy, Urm1 transfer from Uba4 to a substrate involves the RHD in Uba4 (ref. 1). In particular, the Uba4 RHD was suggested to form a transient thioester intermediate

with Urm1, which is then transferred directly to the substrate, bypassing the requirement for a separate E2 (Fig. 4). In *S. cerevisiae*, the cysteine residue in the RHD is required for Urm1-protein ligation<sup>42</sup>.

In 2008, work by Jennifer Schmitz and colleagues<sup>56</sup> showed that the RHD of Uba4 can also form a persulphide and can transfer the terminal sulphur to the C terminus of Urm1 *in vitro*, generating an Urm1-thiocarboxylate. The E1-domain cysteine residue is not required for this, and the authors found no evidence of a Urm1-Uba4 thioester. Instead, they inferred that the adenylated Urm1 is attacked by a persulphide on the RHD, forming an acyl-disulphide intermediate with Urm1, with subsequent release of the thiocarboxylate. They suggested that this could also be an intermediate in Urm1-protein conjugation<sup>56</sup>.

**Figure 3 | General functions of UBL tagging.** **a**, UBL conjugation facilitates protein association by providing an additional binding site. The classic example of this type of regulation is polyubiquitin modification of proteins, which can then bind specific UBD-containing receptors in the proteasome<sup>33</sup>. **b**, UBL conjugation causes a conformational change that enhances binding (as shown) or inhibits binding to a target site. For instance, SUMO attachment to thymine-DNA glycosylase (TDG) triggers a conformational change in TDG that lowers its affinity for DNA<sup>67</sup>. **c**, Modification by one UBL helps recruit a factor that is different from the protein that would be recruited were the substrate modified by another type of UBL. These modifications may be mutually exclusive and can potentially involve the same attachment site. For example, the modification of proliferating cell nuclear antigen (PCNA), a DNA replication and recombination protein, by SUMO, ubiquitin, or ubiquitin polymers causes PCNA to bind distinct factors<sup>68</sup>. **d**, UBL conjugation directly blocks an interaction between two proteins. A potential example of this is the modification of the vaccinia A40R protein by SUMO, which prevents association and aggregation between A40R monomers<sup>69</sup>. Other possible variations on these basic mechanisms are not shown.





A newly discovered function for the Urm1 pathway is in the substitution of S for O at position 2 of the wobble uridine in the anticodon of certain tRNAs, a process that modulates their decoding specificity<sup>58,59</sup>. Urm1 is required for these thiolation reactions, probably functioning as a sulphur carrier in the form of an Urm1–thiocarboxylate. The degree to which tRNA modification by the Urm1 pathway accounts for the pleiotropic physiological roles of Urm1 remains to be determined. Protein modification with Urm1 might also stimulate tRNA thiolation<sup>59</sup>, but a minimal hypothesis would require only the formation of an Urm1–Uba4 acyl disulphide that resolves into the Urm1–thiocarboxylate, from which the sulphur is ultimately transferred to the tRNA.

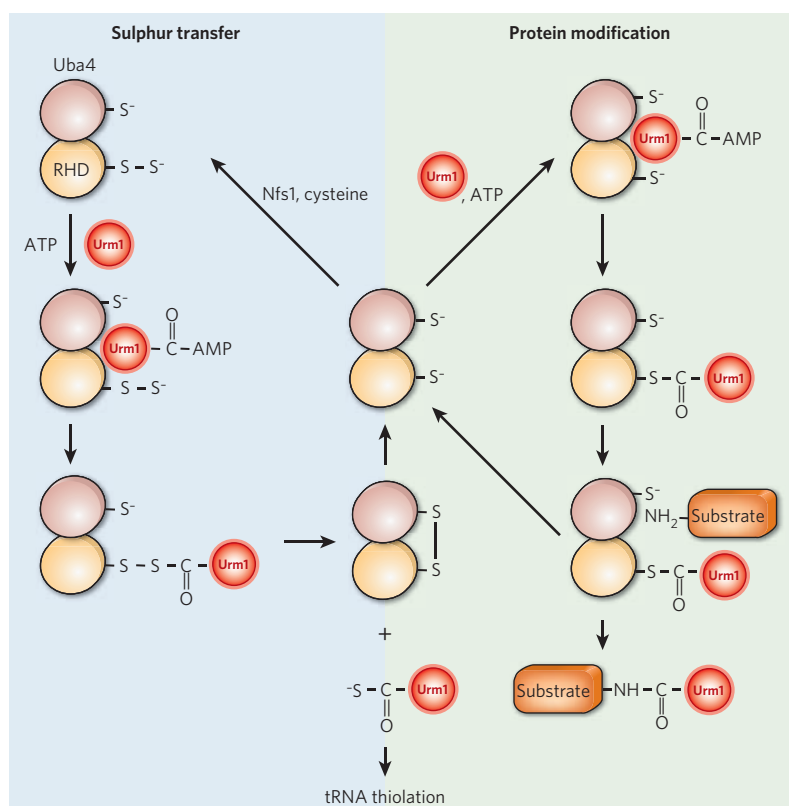
Although the conserved E1-domain cysteine in Uba4 is dispensable for Urm1–thiocarboxylate formation *in vitro*, it is necessary for protein conjugation to Urm1 *in vivo*<sup>54</sup>. This cysteine could function in reductive cleavage of the Urm1–RHD linkage<sup>56</sup>, but this should stimulate formation of an Urm1 C-terminal thiocarboxylate rather than an Urm1–protein amide linkage. Potentially, either the E1-like cysteine undergoes a persulphide exchange with the RHD, freeing the RHD thiol for attack on the Urm1–adenylate (not shown in Fig. 4), or the Urm1–adenylate is directly attacked by a lysine residue in a substrate, although this would not explain the requirement for the two Uba4 active-site cysteine residues. Reductive cleavage of the Urm1–RHD disulphide by the E1-domain active-site sulphydryl group would allow the RHD thiol to be regenerated, potentially making it competent for Urm1–thioester formation. Whether the Urm1–protein ligation mechanism reflects an early precursor to other UBL-conjugation mechanisms is not known. It is conceivable that, during the evolution of UBL-conjugation systems, a distinct E2-like factor could have been co-opted, leading to loss of the RHD from the E1 and thereby eliminating the persulphide–thiocarboxylate ‘side reaction’. Alternatively, eukaryotic UBL modification pathways may have evolved from a distinct prokaryotic  $\beta$ -grasp protein-modification system<sup>6</sup>.

### Radiation of E2s and UBL-specific proteases

When did the E2-like enzymes arise and when did they first associate with E1-like proteins to form the now nearly universal E1–E2 relay used for UBL–protein conjugation? Earlier sequence searches had not revealed any E2-like proteins in bacteria, but recent surveys have revealed an astonishing number of E2-related sequences in the same DNA neighbourhoods (that is, in presumptive operons, domain fusions or co-regulated genes) as UBL-related, E1-like, or JAMM (JAB1/MPN/Mov34) metalloprotease coding sequences<sup>6,60</sup>. In eukaryotes, specific JAMM-class proteases act as deubiquitylating enzymes (DUBs) or UBL-specific proteases (ULPs)<sup>61</sup>.

A striking radiation of E2-like proteins therefore seems to have occurred in bacteria concomitant with the diversification of UBL and E1-like proteins. The subfamily most closely related to classical UBL-conjugating E2 enzymes has been proposed to be ancestral to the eukaryotic UBL E2s<sup>60</sup>. Although none of these prokaryotic E2-like proteins has yet been shown to catalyse UBL–substrate modification in conjunction with an E1, these contextual associations suggest that at least some of them will.

DUBs (or ULPs in the case of UBLs), are often necessary for C-terminal processing of UBL precursors and for removing UBLs from their targets (Box 1). Although multiple JAMM enzymes have now been linked by contextual sequence analysis to potential UBL-modification systems in prokaryotes, some JAMM enzymes participate in sulphur-transfer mechanisms rather than in UBL–protein conjugation. For instance, *Mycobacterium tuberculosis* has an unusual cysteine biosynthetic pathway that involves thiocarboxylate derivatization of CysO, a  $\beta$ -grasp



**Figure 4 | At the crossroads of UBL–protein modification and sulphur transfer.** Two modes of activity for the yeast E1-related protein Uba4 are depicted: in the cycle on the left, Uba4 functions as a sulphurtransferase, transferring sulphur from a persulphide formed on its RHD to the C terminus of Urm1, yielding a Urm1 thiocarboxylate (bottom, middle). The sulphur is ultimately transferred to a specific subset of tRNA. On the right, Uba4 catalyses transfer of Urm1 to protein substrates through a hypothetical Uba4–Urm1 thioester intermediate (bottom right). Although it is possible that the Urm1–Uba4 acyl disulphide (bottom left) functions in Urm1–protein modification, it is difficult to explain the requirement for the E1-like domain cysteine residue in such a scheme. Nfs1 is a cysteine desulphurase that mobilizes sulphur from cysteine<sup>56</sup>.

protein, by the E1-related MoeZ protein<sup>62</sup>. The gene for a JAMM enzyme clusters with the gene for CysO, and it probably hydrolyses cysteine from CysO in the final step of cysteine biosynthesis. Synthesis of the sulphur-containing thioquinolobactin siderophore (an iron-chelating compound) in *Pseudomonas fluorescens* requires a sulphur-carrier protein called QbsE that is related to MoaD and ThiS. QbsE, however, is made in precursor form with two additional amino acids after the diglycine motif<sup>63</sup>. A JAMM protease expressed in the same thioquinolobactin biosynthetic operon cleaves these last two residues from QbsE. Therefore, proteases of the type that eukaryotes use to remove UBLs from protein conjugates might originally have been part of bacterial  $\beta$ -grasp protein-based biosynthetic pathways, just like the UBLs and E1-like (and possibly E2-like) enzymes.

### E1-like activation of a non-UBL substrate

The E1-like superfamily of adenylating enzymes catalyses a spectrum of biochemical reactions that goes beyond C-terminal activation of  $\beta$ -grasp proteins. The best evidence for this comes from enterobacteria that synthesize and secrete the small antibiotic microcin C7 (MccC7). MccC7 is a modified heptapeptide encoded by a large *E. coli* plasmid<sup>64</sup>. An isoasparaginyl moiety at the C terminus of MccC7 has a phosphoramidate linkage to a modified adenylate, and attachment of this modified AMP requires the plasmid-encoded *mccB* product. MccB is a member of the E1-like enzyme superfamily. Therefore, the substrate of this E1-like enzyme is not a  $\beta$ -grasp protein, and the C-terminal modification chemistry is different from that of the sulphurtransferases described above, even though the initial adenylation of a C-terminal  $\alpha$ -carboxylate by the E1-related enzymes is similar.

## Outlook

How much of the diversity of UBL-protein modification or analogous conjugation systems is yet to be discovered? Some UBLs are difficult to recognize by amino-acid sequence comparison, so additional  $\beta$ -grasp/UBL modifiers might still have been overlooked. The exciting possibility of a multitude of prokaryotic UBL-related protein-modification systems, none of which has been analysed experimentally, was made apparent when contextual sequence analyses suggested the existence of a bevy of bacterial regulons that bring together genes encoding novel  $\beta$ -grasp proteins, E1-like enzymes, E2-like proteins and hydrolases related to those in known UBL systems. Moreover, not all E1-like enzymes act on  $\beta$ -grasp/UBL proteins, so further insight into the ability of such enzymes to modify specific proteins or peptides (or other molecules) can be anticipated.

Conversely, there may be intracellular protein–protein conjugation mechanisms that involve neither E1-like adenylating enzymes nor  $\beta$ -grasp proteins, such as intein-mediated protein *trans*-splicing. Along these lines, a 64-residue protein in *M. tuberculosis* called Pup (prokaryotic ubiquitin-like protein) has been shown to modify specific proteins *in vivo* and, remarkably, to target them for degradation by the mycobacterial proteasome<sup>65</sup>. Pup is not a  $\beta$ -grasp/UBL protein, and Pup attachment involves linkage of a substrate lysine to what had originally been a glutamine residue at the Pup C terminus. The terminal Pup glutamine is converted to a glutamate either during or before substrate conjugation. Similar amide-bond-forming reactions are seen with transglutaminases and  $\gamma$ -glutamylcysteine synthetases (involved in glutathione synthesis). In fact, the *M. tuberculosis* PafA protein, which is required for substrate pupylation<sup>65</sup>, has distant sequence similarity to  $\gamma$ -glutamylcysteine synthetases and glutamine synthetases<sup>66</sup>. Mass spectrometry-based proteomic studies could yield further surprises about protein modification and ligation *in vivo*. Viewed through this wider lens, protein–protein conjugation can be seen as a multifaceted and nearly universally employed means of cellular regulation, and we are only just starting to understand some of the underlying mechanisms. ■

- Hochstrasser, M. Evolution and function of ubiquitin-like protein-conjugation systems. *Nature Cell Biol.* **2**, E153–E157 (2000).
- Pickart, C. M. Mechanisms underlying ubiquitination. *Annu. Rev. Biochem.* **70**, 503–533 (2001).
- Xu, P. & Peng, J. Dissecting the ubiquitin pathway by mass spectrometry. *Biochim. Biophys. Acta* **1764**, 1940–1947 (2006).
- Kerscher, O., Felberbaum, R. & Hochstrasser, M. Modification of proteins by ubiquitin and ubiquitin-like proteins. *Annu. Rev. Cell Dev. Biol.* **22**, 159–180 (2006).
- Sharp, P. M. & Li, W.-H. Molecular evolution of ubiquitin genes. *Trends Ecol. Evol.* **2**, 328–332 (1987).
- Iyer, L. M., Burroughs, A. M. & Aravind, L. The prokaryotic antecedents of the ubiquitin-signaling system and the early evolution of ubiquitin-like  $\beta$ -grasp domains. *Genome Biol.* **7**, R60 (2006).
- Highly sensitive sequence comparisons reveal a plethora of prokaryotic UBL/ $\beta$ -grasp proteins and potential UBL-modification pathways.
- Begley, T. P. Cofactor biosynthesis: an organic chemist's treasure trove. *Nature Prod. Rep.* **23**, 15–25 (2006).
- Mukhopadhyay, D. & Riezman, H. Proteasome-independent functions of ubiquitin in endocytosis and signaling. *Science* **315**, 201–205 (2007).
- Schwartz, A. L. & Ciechanover, A. Targeting proteins for destruction by the ubiquitin system: implications for human pathobiology. *Annu. Rev. Pharmacol. Toxicol.* **49**, 73–96 (2009).
- Volker, C. & Lupas, A. N. Molecular evolution of proteasomes. *Curr. Top. Microbiol. Immunol.* **268**, 1–22 (2002).
- Cavaliere-Smith, T. Rooting the tree of life by transition analyses. *Biol. Direct* **1**, 19 (2006).
- Haas, A. L., Ahrens, P., Bright, P. M. & Ankel, H. Interferon induces a 15-kilodalton protein exhibiting marked homology to ubiquitin. *J. Biol. Chem.* **262**, 11315–11323 (1987).
- Loeb, K. R. & Haas, A. L. The interferon-inducible 15-kDa ubiquitin homolog conjugates to intracellular proteins. *J. Biol. Chem.* **267**, 7806–7813 (1992).
- Yuan, W. & Krug, R. M. Influenza B virus NS1 protein inhibits conjugation of the interferon (IFN)-induced ubiquitin-like ISG15 protein. *EMBO J.* **20**, 362–371 (2001).
- Zhao, C. *et al.* The UbcH8 ubiquitin E2 enzyme is also the E2 enzyme for ISG15, an IFN- $\alpha$ / $\beta$ -induced ubiquitin-like protein. *Proc. Natl Acad. Sci. USA* **101**, 7578–7582 (2004).
- Kim, K. I., Giannakopoulos, N. V., Virgin, H. W. & Zhang, D. E. Interferon-inducible ubiquitin E2, Ubc8, is a conjugating enzyme for protein ISGylation. *Mol. Cell. Biol.* **24**, 9592–9600 (2004).
- Durfee, L. A., Kelley, M. L. & Huibregtse, J. M. The basis for selective E1–E2 interactions in the ISG15 conjugation system. *J. Biol. Chem.* **283**, 23895–23902 (2008).
- Lenschow, D. J. *et al.* IFN-stimulated gene 15 functions as a critical antiviral molecule against influenza, herpes, and Sindbis viruses. *Proc. Natl Acad. Sci. USA* **104**, 1371–1376 (2007).
- Okumura, A., Pitha, P. M. & Harty, R. N. ISG15 inhibits Ebola VP40 VLP budding in an L-domain-dependent manner by blocking Nedd4 ligase activity. *Proc. Natl Acad. Sci. USA* **105**, 3974–3979 (2008).
- Malakhova, O. A. & Zhang, D. E. ISG15 inhibits Nedd4 ubiquitin E3 activity and enhances the innate antiviral response. *J. Biol. Chem.* **283**, 8783–8787 (2008).
- Peng, J. *et al.* A proteomics approach to understanding protein ubiquitination. *Nature Biotechnol.* **21**, 921–926 (2003).
- Ohsumi, Y. Molecular dissection of autophagy: two ubiquitin-like systems. *Nature Rev. Mol. Cell Biol.* **2**, 211–216 (2001).
- Liu, Y., Fallon, L., Lashuel, H. A., Liu, Z. & Lansbury, P. T. Jr. The *UCH-L1* gene encodes two opposing enzymatic activities that affect  $\alpha$ -synuclein degradation and Parkinson's disease susceptibility. *Cell* **111**, 209–218 (2002).
- Dassa, B., Yanai, I. & Pietrokovski, S. New type of polyubiquitin-like genes with intein-like autoprocessing domains. *Trends Genet.* **20**, 538–542 (2004).
- Huang, T. T. *et al.* Regulation of monoubiquitinated PCNA by DUB autocleavage. *Nature Cell Biol.* **8**, 339–347 (2006).
- Burroughs, A. M., Balaji, S., Iyer, L. M. & Aravind, L. Small but versatile: the extraordinary functional and structural diversity of the  $\beta$ -grasp fold. *Biol. Direct* **2**, 18 (2007).
- Hershko, A. & Ciechanover, A. The ubiquitin system for protein degradation. *Annu. Rev. Biochem.* **61**, 761–807 (1992).
- Goldberg, A. L. & Rock, K. L. Proteolysis, proteasomes and antigen presentation. *Nature* **357**, 375–379 (1992).
- Hochstrasser, M. Ubiquitin-dependent protein degradation. *Annu. Rev. Genet.* **30**, 405–439 (1996).
- Thrower, J. S., Hoffman, L., Rechsteiner, M. & Pickart, C. M. Recognition of the polyubiquitin proteolytic signal. *EMBO J.* **19**, 94–102 (2000).
- Deveraux, Q., Ustrell, V., Pickart, C. & Rechsteiner, M. A 26 S protease subunit that binds ubiquitin conjugates. *J. Biol. Chem.* **269**, 7059–7061 (1994).
- Lam, Y. A., Lawson, T. G., Velayutham, M., Zweier, J. L. & Pickart, C. M. A proteasomal ATPase subunit recognizes the polyubiquitin degradation signal. *Nature* **416**, 763–767 (2002).
- Husnjak, K. *et al.* Proteasome subunit Rpn13 is a novel ubiquitin receptor. *Nature* **453**, 481–488 (2008).
- Chen, L. & Madura, K. Rad23 promotes the targeting of proteolytic substrates to the proteasome. *Mol. Cell. Biol.* **22**, 4902–4913 (2002).
- Funakoshi, M., Sasaki, T., Nishimoto, T. & Kobayashi, H. Budding yeast Dsk2p is a polyubiquitin-binding protein that can interact with the proteasome. *Proc. Natl Acad. Sci. USA* **99**, 745–750 (2002).
- Kim, I., Mi, K. & Rao, H. Multiple interactions of rad23 suggest a mechanism for ubiquitylated substrate delivery important in proteolysis. *Mol. Biol. Cell* **15**, 3357–3365 (2004).
- Hurley, J. H., Lee, S. & Prag, G. Ubiquitin-binding domains. *Biochem. J.* **399**, 361–372 (2006).
- Varadan, R. *et al.* Solution conformation of Lys63-linked di-ubiquitin chain provides clues to functional diversity of polyubiquitin signaling. *J. Biol. Chem.* **279**, 7055–7063 (2004).
- Kerscher, O. SUMO junction — what's your function? New insights through SUMO-interacting motifs. *EMBO Rep.* **8**, 550–555 (2007).
- Song, J., Zhang, Z., Hu, W. & Chen, Y. Small ubiquitin-like modifier (SUMO) recognition of a SUMO binding motif: a reversal of the bound orientation. *J. Biol. Chem.* **280**, 40122–40129 (2005).
- Reverter, D. & Lima, C. D. Insights into E3 ligase activity revealed by a SUMO–RanGAP1–Ubc9–Nup358 complex. *Nature* **435**, 687–692 (2005).
- Hochstrasser, M. in *Protein Degradation: The Ubiquitin-Proteasome System* (eds Mayer, R. J., Ciechanover, A. & Rechsteiner, M.) 249–278 (Wiley, 2006).
- Archer, C. T. *et al.* Physical and functional interactions of monoubiquitylated transactivators with the proteasome. *J. Biol. Chem.* **283**, 21789–21798 (2008).
- Mahajan, R., Delphin, C., Guan, T., Gerace, L. & Melchior, F. A small ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore complex protein RanBP2. *Cell* **88**, 97–107 (1997).
- McGrath, J. P., Jentsch, S. & Varshavsky, A. UBA1: an essential yeast gene encoding ubiquitin-activating enzyme. *EMBO J.* **10**, 227–236 (1991).
- Rajagopalan, K. V. Biosynthesis and processing of the molybdenum cofactors. *Biochem. Soc. Trans.* **25**, 757–761 (1997).
- Taylor, S. V. *et al.* Thiamin biosynthesis in *Escherichia coli*. Identification of this thiocarboxylate as the immediate sulfur donor in the thiazole formation. *J. Biol. Chem.* **273**, 16555–16560 (1998).
- Appleyard, M. V. *et al.* The *Aspergillus nidulans* *cnx*F gene and its involvement in molybdopterin biosynthesis. Molecular characterization and analysis of *in vivo* generated mutants. *J. Biol. Chem.* **273**, 14869–14876 (1998).
- Leimkuhler, S., Wuebbens, M. M. & Rajagopalan, K. V. Characterization of *Escherichia coli* MoeB and its involvement in the activation of molybdopterin synthase for the biosynthesis of the molybdenum cofactor. *J. Biol. Chem.* **276**, 34695–34701 (2001).
- Rudolph, M. J., Wuebbens, M. M., Rajagopalan, K. V. & Schindelin, H. Crystal structure of molybdopterin synthase and its evolutionary relationship to ubiquitin activation. *Nature Struct. Biol.* **8**, 42–46 (2001).
- Wang, C., Xi, J., Begley, T. P. & Nicholson, L. K. Solution structure of ThiS and implications for the evolutionary roots of ubiquitin. *Nature Struct. Biol.* **8**, 47–51 (2001).
- Huang, D. T., Walden, H., Duda, D. & Schulman, B. A. Ubiquitin-like protein activation. *Oncogene* **23**, 1958–1971 (2004).
- Duda, D. M., Walden, H., Sfondouris, J. & Schulman, B. A. Structural analysis of *Escherichia coli* ThiF. *J. Mol. Biol.* **349**, 774–786 (2005).
- Furukawa, K., Mizushima, N., Noda, T. & Ohsumi, Y. A protein conjugation system in yeast with homology to biosynthetic enzyme reaction of prokaryotes. *J. Biol. Chem.* **275**, 7462–7465 (2000).
- Goehring, A. S., Rivers, D. M. & Sprague, G. F. Jr. Attachment of the ubiquitin-related protein Urm1p to the antioxidant protein Ahp1p. *Eukaryot. Cell* **2**, 930–936 (2003).



56. Schmitz, J. *et al.* The sulfurtransferase activity of Uba4 presents a link between ubiquitin-like protein conjugation and activation of sulfur carrier proteins. *Biochemistry* **47**, 6479–6489 (2008).  
**Identifies a persulphide in Uba4 and formation of an Urm1 thiocarboxylate, suggesting a dual function in protein conjugation and sulphur transfer.**
57. Mueller, E. G. Trafficking in persulfides: delivering sulfur in biosynthetic pathways. *Nature Chem. Biol.* **2**, 185–194 (2006).
58. Nakai, Y., Nakai, M. & Hayashi, H. Thio-modification of yeast cytosolic tRNA requires a ubiquitin-related system that resembles bacterial sulfur transfer systems. *J. Biol. Chem.* **283**, 27469–27476 (2008).  
**This study and ref. 58 implicate the Urm1 thiocarboxylate as a potential sulphur carrier in selective tRNA thiolation.**
60. Burroughs, A. M., Jaffee, M., Iyer, L. M. & Aravind, L. Anatomy of the E2 ligase fold: implications for enzymology and evolution of ubiquitin/Ub-like protein conjugation. *J. Struct. Biol.* **162**, 205–218 (2008).
61. Amerik, A. Y. & Hochstrasser, M. Mechanism and function of deubiquitinating enzymes. *Biochim. Biophys. Acta* **1695**, 189–207 (2004).
62. Burns, K. E. *et al.* Reconstitution of a new cysteine biosynthetic pathway in *Mycobacterium tuberculosis*. *J. Am. Chem. Soc.* **127**, 11602–11603 (2005).
63. Godert, A. M., Jin, M., McLafferty, F. W. & Begley, T. P. Biosynthesis of the thioquinolobactin siderophore: an interesting variation on sulfur transfer. *J. Bacteriol.* **189**, 2941–2944 (2007).
64. Roush, R. F., Nolan, E. M., Lohr, F. & Walsh, C. T. Maturation of an *Escherichia coli* ribosomal peptide antibiotic by ATP-consuming N–P bond formation in microcin C7. *J. Am. Chem. Soc.* **130**, 3603–3609 (2008).  
**Describes the role of an E1-like adenylating enzyme in the activation of a non- $\beta$ -grasp peptide.**
65. Pearce, M. J., Mintseris, J., Ferreyra, J., Gygi, S. P. & Darwin, K. H. Ubiquitin-like protein involved in the proteasome pathway of *Mycobacterium tuberculosis*. *Science* **322**, 1104–1107 (2008).  
**Identifies a non- $\beta$ -grasp protein in *M. tuberculosis* that modifies selected protein substrates and targets them for proteasomal degradation.**
66. Iyer, L. M., Burroughs, A. M. & Aravind, L. Unraveling the biochemistry and provenance of pupylation: a prokaryotic analog of ubiquitination. *Biol. Direct* **3**, 45 (2008).
67. Steinacher, R. & Schar, P. Functionality of human thymine DNA glycosylase requires SUMO-regulated changes in protein conformation. *Curr. Biol.* **15**, 616–623 (2005).
68. Ulrich, H. D. How to activate a damage-tolerant polymerase: consequences of PCNA modifications by ubiquitin and SUMO. *Cell Cycle* **3**, 15–18 (2004).
69. Palacios, S. *et al.* Quantitative SUMO-1 modification of a vaccinia virus protein is required for its specific localization and prevents its self-association. *Mol. Biol. Cell* **16**, 2822–2835 (2005).
70. Ciechanover, A. & Ben-Saadon, R. N-terminal ubiquitination: more protein substrates join in. *Trends Cell Biol.* **14**, 103–106 (2004).
71. Cadwell, K. & Coscoy, L. Ubiquitination on nonlysine residues by a viral E3 ubiquitin ligase. *Science* **309**, 127–130 (2005).
72. Ravid, T. & Hochstrasser, M. Autoregulation of an E2 enzyme by ubiquitin-chain assembly on its catalytic residue. *Nature Cell Biol.* **9**, 422–427 (2007).
73. Wang, X. *et al.* Ubiquitination of serine, threonine, or lysine residues on the cytoplasmic tail can induce ERAD of MHC-I by viral E3 ligase mK3. *J. Cell Biol.* **177**, 613–624 (2007).
74. Lee, I. & Schindelin, H. Structural insights into E1-catalyzed ubiquitin activation and transfer to conjugating enzymes. *Cell* **134**, 268–278 (2008).
75. Nijman, S. M. *et al.* A genomic and functional inventory of deubiquitinating enzymes. *Cell* **123**, 773–786 (2005).

**Acknowledgements** I thank V. J. Rubenstein, A. Kusmierczyk and J. Bloom for comments on the manuscript. Work in my laboratory is funded by grants from the National Institutes of Health.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The author declares no competing financial interests. Correspondence should be addressed to the author (mark.hochstrasser@yale.edu).

# Ubiquitylation in innate and adaptive immunity

Vijay G. Bhoj<sup>1</sup> & Zhijian J. Chen<sup>1,2</sup>

**Protein ubiquitylation has emerged as a key mechanism that regulates immune responses. Much like phosphorylation, ubiquitylation is a reversible covalent modification that regulates the stability, activity and localization of target proteins. As such, ubiquitylation regulates the development of the immune system and many phases of the immune response, including its initiation, propagation and termination. Recent work has shown that several ubiquitin ligases help to prevent the immune system from attacking self tissues. The dysfunction of several ubiquitin ligases has been linked to autoimmune diseases.**

A safe and effective immune system must be capable of eradicating or containing a vast variety of invading microorganisms but must avoid attacking self tissues. To accomplish such a daunting task, the immune system is under stringent regulation. As one of the most important regulatory mechanisms in biology, ubiquitylation is extensively used to orchestrate an appropriate immune response. Early insights into the role of ubiquitin in immunity were provided by the study of the nuclear factor- $\kappa$ B (NF- $\kappa$ B) pathway, and research in the past few years has demonstrated that ubiquitin has a central role in several pathways leading to NF- $\kappa$ B activation. NF- $\kappa$ B is key to both innate and adaptive immunity, and its regulation provides a model for understanding the role of ubiquitin in the immune system.

Ubiquitin is a highly conserved 76-amino-acid polypeptide that is covalently attached to one or more lysine residues of cellular proteins through an enzymatic cascade involving three classes of enzyme termed E1, E2 and E3. Ubiquitin itself contains seven lysines, and each of these can be further conjugated by the carboxyl terminus of another ubiquitin to form a polyubiquitin chain. It is believed that polyubiquitin chains linked through lysine at position 48 of ubiquitin (Lys 48) target protein substrates for degradation by the proteasome, whereas polyubiquitin chains of alternative linkages (such as Lys 63) carry out signalling functions independent of proteolysis<sup>1</sup>. In addition, some proteins such as the histones H2A and H2B are modified by a single ubiquitin that has a non-degradative regulatory function.

Like phosphorylation, ubiquitylation can be reversed, in this case by deubiquitylation enzymes (DUBs), making it ideally suited for regulatory controls. Ubiquitin or polyubiquitin chains serve as three-dimensional signals that are recognized by many different types of ubiquitin-binding domain (UBD). The human proteome contains two ubiquitin E1 enzymes, about 50 E2s, 600 E3s, 90 DUBs and 20 types of UBD, underscoring the central role of ubiquitin in cell regulation. E3s are the main determinant of substrate specificity and are categorized into two classes, one containing a HECT domain and the other containing a RING or RING-like (for example, U-box or PHD) domain. Both HECT- and RING-domain E3s have key roles in regulating immune responses.

Here we consider the role of the ubiquitin system in NF- $\kappa$ B regulation. We then discuss recent advances in understanding the role of ubiquitin in innate and adaptive immunity. Finally, we discuss several ubiquitin ligases that function to prevent autoimmunity. Space limitations prevent us from discussing the role of ubiquitin in antigen presentation and

pathogens' hijacking of the ubiquitin system to evade the host immune system. For these topics, the readers are referred to two excellent recent reviews<sup>2,3</sup>.

## Ubiquitin in NF- $\kappa$ B regulation

Since its discovery about two decades ago, NF- $\kappa$ B has been shown to have a pivotal role in many biological processes, including inflammation, immunity and cell survival. The NF- $\kappa$ B pathway is also an excellent model system for studying cell signalling, as the activity of NF- $\kappa$ B is tightly regulated. Although widely expressed, NF- $\kappa$ B is kept inactive in most cells by being sequestered in the cytoplasm by members of the I $\kappa$ B family of inhibitory proteins. In response to stimulation by a large variety of agents, including many derived from microorganisms, I $\kappa$ Bs are rapidly degraded by the ubiquitin-proteasome pathway, allowing NF- $\kappa$ B to enter the nucleus to turn on a wide spectrum of genes.

The NF- $\kappa$ B family consists of five members: p65 (REL-A), c-REL, REL-B, p50 and p52. All contain a REL-homology domain (RHD) that is responsible for DNA binding, dimerization, nuclear translocation and I $\kappa$ B binding. In addition, p65, c-REL and REL-B contain a trans-activation domain (TAD) that is required for gene activation. Both p50 and p52 are processed from precursors, p105 and p100, respectively, and must dimerize with a TAD-containing member to activate gene transcription. Both p105 and p100 contain C-terminal ankyrin repeats, which are also found in I $\kappa$ B proteins, including I $\kappa$ B- $\alpha$ , I $\kappa$ B- $\beta$  and I $\kappa$ B- $\epsilon$ . The ankyrin repeats bind to the RHD of NF- $\kappa$ B, masking its nuclear localization sequence, thereby retaining NF- $\kappa$ B in the cytoplasm.

## Degradation of I $\kappa$ Bs and processing of NF- $\kappa$ B precursors

The activation of NF- $\kappa$ B is broadly classified into one of two schemes, canonical and non-canonical<sup>4</sup>. In the canonical pathway, which operates in most cells, stimulation of cells leads to the rapid phosphorylation of I $\kappa$ B by a large kinase complex called I $\kappa$ B kinase (IKK), which consists of two catalytic subunits, IKK- $\alpha$  and IKK- $\beta$ , and an essential regulatory subunit termed NEMO, also known as IKK- $\gamma$  (Fig. 1). Phosphorylated I $\kappa$ B is then recruited to a ubiquitin ligase complex through the F-box protein  $\beta$ TrCP, which contains a WD40-repeat domain that binds specifically to I $\kappa$ B that has two phosphorylated serine residues at its amino terminus. Through its F-box,  $\beta$ TrCP binds to SKP1, an adaptor protein involved in the assembly of the large ubiquitin E3 complex SCF, which also contains CUL1 and ROC1 (also known as Rbx1). ROC1 is a RING-domain protein that recruits an E2 to catalyse

<sup>1</sup>Department of Molecular Biology and <sup>2</sup>Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA.



the polyubiquitylation of phosphorylated I $\kappa$ B. Polyubiquitylated I $\kappa$ B remains bound to NF- $\kappa$ B but is selectively degraded by the 26S proteasome, thereby liberating NF- $\kappa$ B.

The non-canonical pathway of NF- $\kappa$ B activation, which operates mainly in B cells, involves the proteolytic processing of the NF- $\kappa$ B precursor p100 to the mature p52 subunit. Stimulation of a subset of members of the tumour necrosis factor receptor (TNFR) superfamily, including the CD40 and BAFF receptors, results in activation of the protein kinase NIK, which in turn phosphorylates and activates IKK- $\alpha$ . IKK- $\alpha$  phosphorylates p100 at two serine residues in the C terminus, which allows p100 to be recognized by the SCF- $\beta$ TrCP E3 complex. The polyubiquitylation of p100, however, does not result in its complete degradation by the proteasome. Instead, the proteasome degrades only the C terminus of p100, which contains the ankyrin repeats, leaving the N-terminal RHD domain (p52) intact. Then p52 dimerizes with its partner REL-B to induce a gene-expression program that is important for B-cell maturation and activation. The ubiquitin-proteasome pathway also processes p105 to p50 (ref. 5); this is largely a constitutive process, although some stimuli such as phorbol ester have been reported to enhance p105 processing.

### Ubiquitin-mediated activation of protein kinases

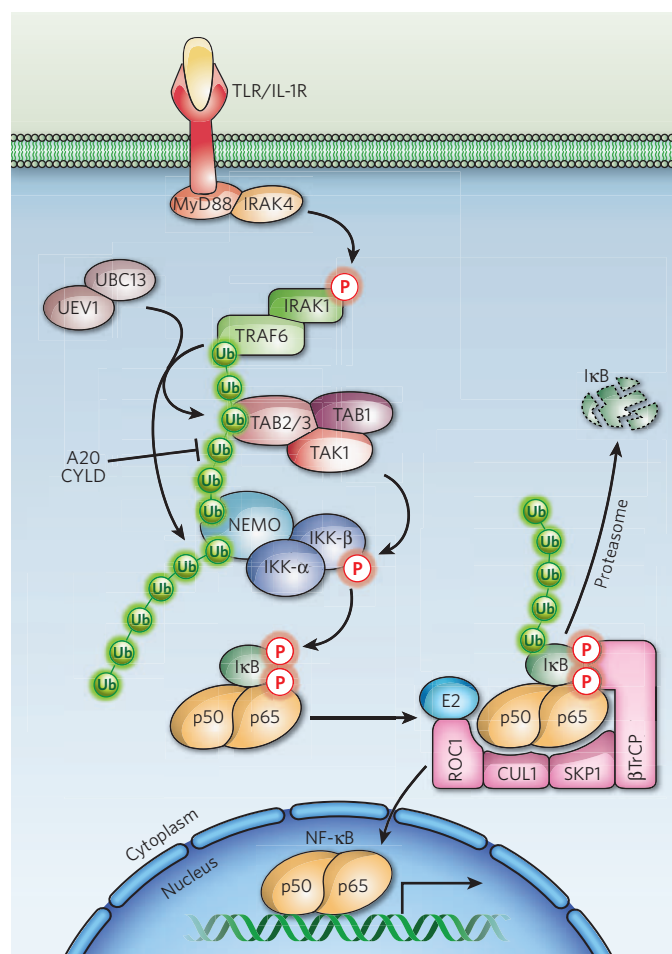
Phosphorylation of I $\kappa$ Bs and p100 by IKK is a prerequisite for their ubiquitylation and subsequent degradation or processing by the proteasome. It is therefore critical to understand how IKK is regulated by upstream signals. Surprisingly, ubiquitylation has a major role in IKK activation through a mechanism independent of proteasomal degradation. The first clue to the role of ubiquitin in IKK activation was provided in a biochemical fractionation experiment aimed at identifying the I $\kappa$ B kinase<sup>6</sup>. It was found that this kinase activity was greatly stimulated by polyubiquitylation *in vitro*, and that this activation occurs in the absence of the proteasome or in the presence of proteasome inhibitors.

Further insight into the mechanism of the ubiquitin-mediated activation of IKK was provided by biochemical studies of TRAF6, a RING-domain protein essential for IKK activation by interleukin-1 (IL-1) and Toll-like receptors (TLRs) (Fig. 1). TRAF6 functions as a ubiquitin E3, which, in conjunction with a ubiquitin E2 complex composed of UBC13 and UEV1A, catalyses the synthesis of a polyubiquitin chain linked through Lys 63 of ubiquitin<sup>7</sup>. This polyubiquitylation activates a protein kinase complex comprising TAK1, TAB1 and TAB2 (or TAB3)<sup>8</sup>. TAB2 and TAB3 bind preferentially to Lys 63 polyubiquitin chains<sup>9</sup>. This binding leads to the activation of TAK1, perhaps by triggering its autophosphorylation. Activated TAK1 phosphorylates IKK- $\beta$  at two serine residues within its activation loop, leading to IKK activation. TAK1 also phosphorylates the MAPK kinases, such as MKK6 and MKK7, leading to activation of the JNK and p38 kinase pathways.

Phosphorylation of IKK- $\beta$  by TAK1 requires NEMO, which contains a unique ubiquitin-binding domain<sup>10,11</sup>. Point mutations that disrupt the ubiquitin binding of NEMO, such as those found in ectodermal dysplasia with immunodeficiency, abrogate IKK activation. NEMO is also polyubiquitylated at various sites following the stimulation of cells with different NF- $\kappa$ B activators, and this may contribute to IKK activation<sup>12</sup>. Other proteins in the IL-1 pathway known to be modified by Lys 63 polyubiquitin chains include IRAK1 and TRAF6, but further research is required to determine whether the ubiquitylation of any of these proteins is important for IKK activation.

### DUBs prevent protein kinase activation

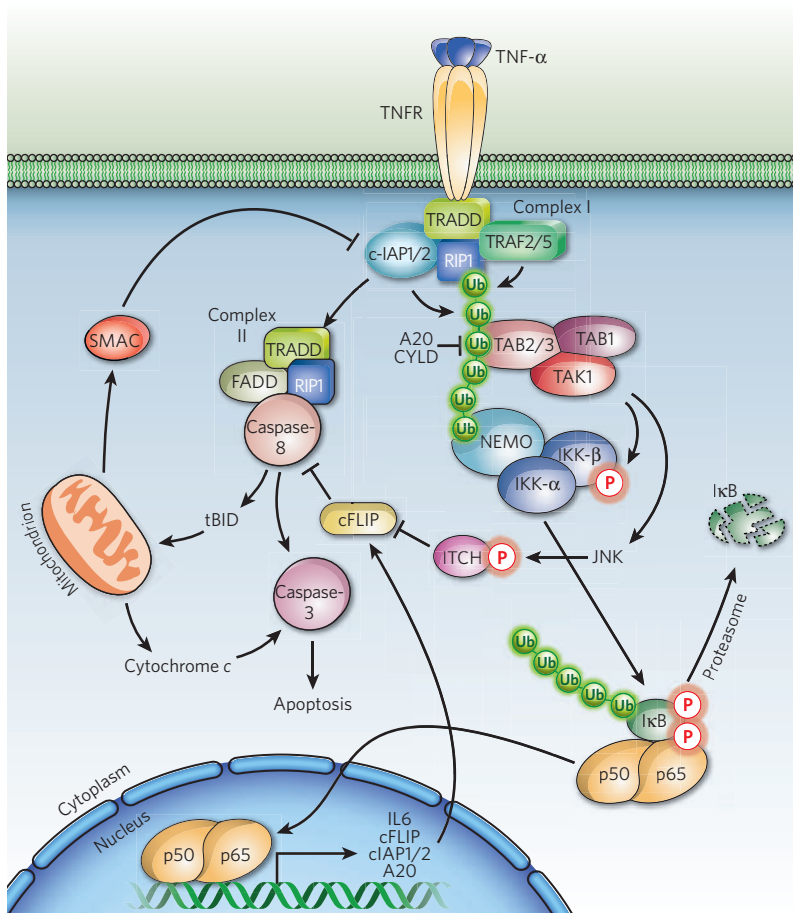
Further support for the idea that polyubiquitylation is important for IKK activation comes from the discovery of several DUBs that negatively regulate NF- $\kappa$ B upstream of IKK<sup>13</sup>. These enzymes include CYLD, a tumour suppressor implicated in several human diseases, including cylindromatosis and multiple myeloma, and A20, a potent inhibitor of NF- $\kappa$ B. CYLD contains three cytoskeleton-associated protein (CAP)-Gly domains followed by a C-terminal ubiquitin-specific protease (USP) catalytic domain<sup>14</sup>. CYLD has a high specificity in cleaving Lys-63-linked polyubiquitin chains, and thereby prevents the activation of TAK1, JNK



**Figure 1 | Ubiquitin-mediated activation of NF- $\kappa$ B in the TLR and IL-1R pathways.** Ligand-bound TLRs and IL-1R recruit the adaptor MyD88, which in turn recruits the kinases IRAK1 and IRAK4. IRAK4 is activated and phosphorylates IRAK1. Phosphorylated IRAK1 then associates with TRAF6. This association activates the ligase function of TRAF6, which, with the UBC13/UEV1 E2 complex, catalyses the synthesis of Lys-63-linked polyubiquitin chains. These chains serve as a scaffold to recruit both the TAK1 and IKK complexes through their respective ubiquitin-binding subunits, TAB2/3 and NEMO. As a result of their proximity, TAK1 can phosphorylate IKK- $\beta$ , which then phosphorylates I $\kappa$ B. Then  $\beta$ TrCP, a component of the SCF E3 complex, binds specifically to the phosphorylated form of I $\kappa$ B and, with an E2, catalyses the polyubiquitylation of I $\kappa$ B. I $\kappa$ B is then degraded by the proteasome, allowing NF- $\kappa$ B to enter the nucleus to begin transcriptional regulation.

and IKK in several signalling pathways (Fig. 1). Several mouse models with CYLD mutations have been generated in different laboratories. Although the reported phenotypes of these mice are not identical, excessive inflammation associated with hyperactivation of IKK has consistently been observed<sup>15</sup>.

A20 contains an N-terminal ovarian tumour (OTU)-type DUB domain, followed by seven zinc-finger domains at the C terminus<sup>15,16</sup>. It has been proposed that the OTU domain of A20 catalyses the disassembly of Lys-63-linked polyubiquitin chains from target proteins such as RIP1, a key signalling protein required for IKK activation by TNF- $\alpha$ <sup>17</sup> (Fig. 2). Furthermore, one of the C-terminal zinc-finger domains has ubiquitin ligase activity and catalyses the Lys-48-linked polyubiquitylation of RIP1, thereby targeting RIP1 for degradation by the proteasome. However, A20 cleaves both Lys 48 and Lys 63 linkages *in vitro*<sup>18</sup>. It is possible that other cellular proteins assist A20 in achieving substrate and ubiquitin-linkage specificity *in vivo*. Indeed, A20 associates with ABINs (A20-binding inhibitors of NF- $\kappa$ B) and TAX1BP1 (TAX1-binding protein 1), all of which contain ubiquitin-binding domains<sup>19</sup>. In cells that lack TAX1BP1, A20 fails to remove polyubiquitin chains from



**Figure 2 | Ubiquitin coordinates NF- $\kappa$ B activation and apoptosis triggered by TNFR.** After TNF- $\alpha$  binding, the trimeric TNFR binds to TRADD, which recruits TRAF2/5, RIP1 and cIAP1/2 to form complex I. Within this complex, RIP1 is polyubiquitinated, which recruits and activates the TAK1 and IKK complexes. Subsequent to complex I formation and activation of the NF- $\kappa$ B signalling cascade, a cytoplasmic complex II containing TRADD, FADD, RIP1 and procaspase-8 is formed. Procaspase-8 is processed to caspase-8, which cleaves procaspase-3 into the mature effector caspase-3, to initiate apoptosis. Caspase-8 also processes BID into a truncated form, tBID, which inserts into the mitochondrial membrane causing mitochondrial damage. This causes the mitochondrion to release cytochrome c, which contributes to further caspase-3 activation, and SMAC. SMAC triggers the degradation of cIAPs, which normally prevent RIP1 from forming complex II. Most cells stimulated with TNF- $\alpha$  do not undergo apoptosis, as the induction of cFLIP by NF- $\kappa$ B inhibits caspase-8 activity. Under certain conditions, activated JNK phosphorylates the E3 ligase ITCH, which leads to cFLIP degradation, again promoting apoptosis.

TRAF6 and RIP1, resulting in hyperactivation of NF- $\kappa$ B. A20 also binds to ITCH, a HECT-domain E3 known to have an important role in preventing autoimmunity<sup>20</sup>. ITCH is required for the TNF $\alpha$ -induced degradation of RIP1, raising the question of whether ITCH or A20, or both, is directly responsible for RIP1 ubiquitylation and degradation. In any case, a key mechanism of NF- $\kappa$ B inhibition by A20 is likely through the formation of a large ubiquitin-editing complex.

### Ubiquitin in innate immunity

Innate immunity is an ancient but highly effective first line of defence against microorganisms. When a pathogen breaches our physical barriers, it is detected by various immune as well as non-immune cells. Detection is mediated by receptors that bind to pathogen-associated molecular patterns (PAMPs), elements of pathogens that are essential for their survival and are typically distinct from host molecules<sup>21</sup>. The binding of PAMPs to host receptors, collectively termed pattern recognition receptors (PRR), activates signalling pathways leading to the induction of various antimicrobial molecules such as cytokines and chemokines. Some of these molecules, such as TNF- $\alpha$ , trigger powerful inflammatory responses that limit pathogen growth and recruit immune cells to the site of infection. Others, such as type I interferons (IFN-I), not only activate signalling cascades that inhibit viral replication and assembly, but are also required for the activation of the adaptive immune system to eradicate the infection.

Unlike antigen receptors expressed on T and B cells, which are generated by somatic gene rearrangement and hypermutation, PRRs are germline-encoded receptors that are evolutionarily conserved. Three classes of PRR have been extensively studied: TLRs, NOD-like receptors (NLRs) and RIG-I-like receptors (RLRs). All these receptors activate signalling cascades that lead to the activation of MAP kinases and NF- $\kappa$ B. In addition, RLRs and some TLRs activate interferon regulatory factors (IRFs), which function together with NF- $\kappa$ B and other transcription factors to induce production of IFN-I and other effector molecules.

Ubiquitylation has a crucial role in signalling by all three classes of PRR (Fig. 3). Furthermore, intensive research into TNF signalling pathways has shown that ubiquitylation controls a cell's life and death through coordinated regulation of NF- $\kappa$ B and apoptosis (Fig. 2).

### TNF signalling

As one of the major effector molecules of the innate immune responses triggered by all PRRs, TNF- $\alpha$  has been a subject of intense research. It exerts its biological functions primarily through its receptor TNFR1 and, to a lesser extent, TNFR2. TNFR1 and TNFR2 are the founding members of TNFR superfamily. Stimulation of some of these receptors, including TNFR1, can also lead to cell death. Thus the TNFR1 pathway serves as a model system for understanding immune signalling as well as cell death.

The binding of TNF- $\alpha$ , which is a trimer, leads to the trimerization of TNFR1. TNFR1 recruits a signalling complex (complex I) that includes TRADD, RIP1, TRAF2, and TRAF5, and cIAP1 and cIAP2 (ref. 22) (Fig. 2). Notably, the TRAF and cIAP proteins in this complex are RING-domain ubiquitin ligases. In addition, RIP1 is polyubiquitinated at a specific site following TNF- $\alpha$  stimulation, and a mutation that prevents RIP1 ubiquitylation also blocks IKK activation<sup>10,23</sup>. Genetic studies suggest that TRAF2 is required for RIP1 polyubiquitylation, but there is no direct evidence that TRAF2 or TRAF5 catalyses the ubiquitylation of RIP1 *in vitro*<sup>24</sup>. Recent studies suggest that cIAP1 and cIAP2 can catalyse the polyubiquitylation of RIP1 and are required for IKK activation by TNF- $\alpha$ <sup>25</sup>. Polyubiquitylated RIP1 recruits the TAK1 kinase complex through the binding between the polyubiquitin chains and TAB2 or TAB3 (ref. 9). The polyubiquitin chains also bind to NEMO, thereby recruiting IKK to TAK1 in the membrane receptor complex, leading to the phosphorylation of IKK by TAK1 (refs 10, 11).

After the formation of the TNFR complex I, the death-domain proteins TRADD and RIP1 dissociate from the membrane receptor complex and associate with FADD to form complex II in the cytosol<sup>22</sup> (Fig. 2).



FADD facilitates the autocatalytic processing of procaspase-8 to mature caspase-8, which in turn cleaves procaspase-3 to initiate the apoptotic program. However, apoptosis induced by TNF- $\alpha$  does not normally occur because NF- $\kappa$ B turns on the production of antiapoptotic proteins such as c-FLIP, a potent inhibitor of caspase-8. To allow apoptosis to proceed, c-FLIP must be removed, either by preventing its synthesis or promoting its degradation. The synthesis of c-FLIP and other apoptosis inhibitors can be blocked with protein-synthesis inhibitors such as cyclohexamide, or more specifically by inhibiting the NF- $\kappa$ B pathway; this shifts the balance toward TNF $\alpha$ -induced apoptosis. The degradation of c-FLIP is mediated by ITCH, a ubiquitin ligase whose activity is enhanced by JNK-catalysed phosphorylation<sup>26</sup>.

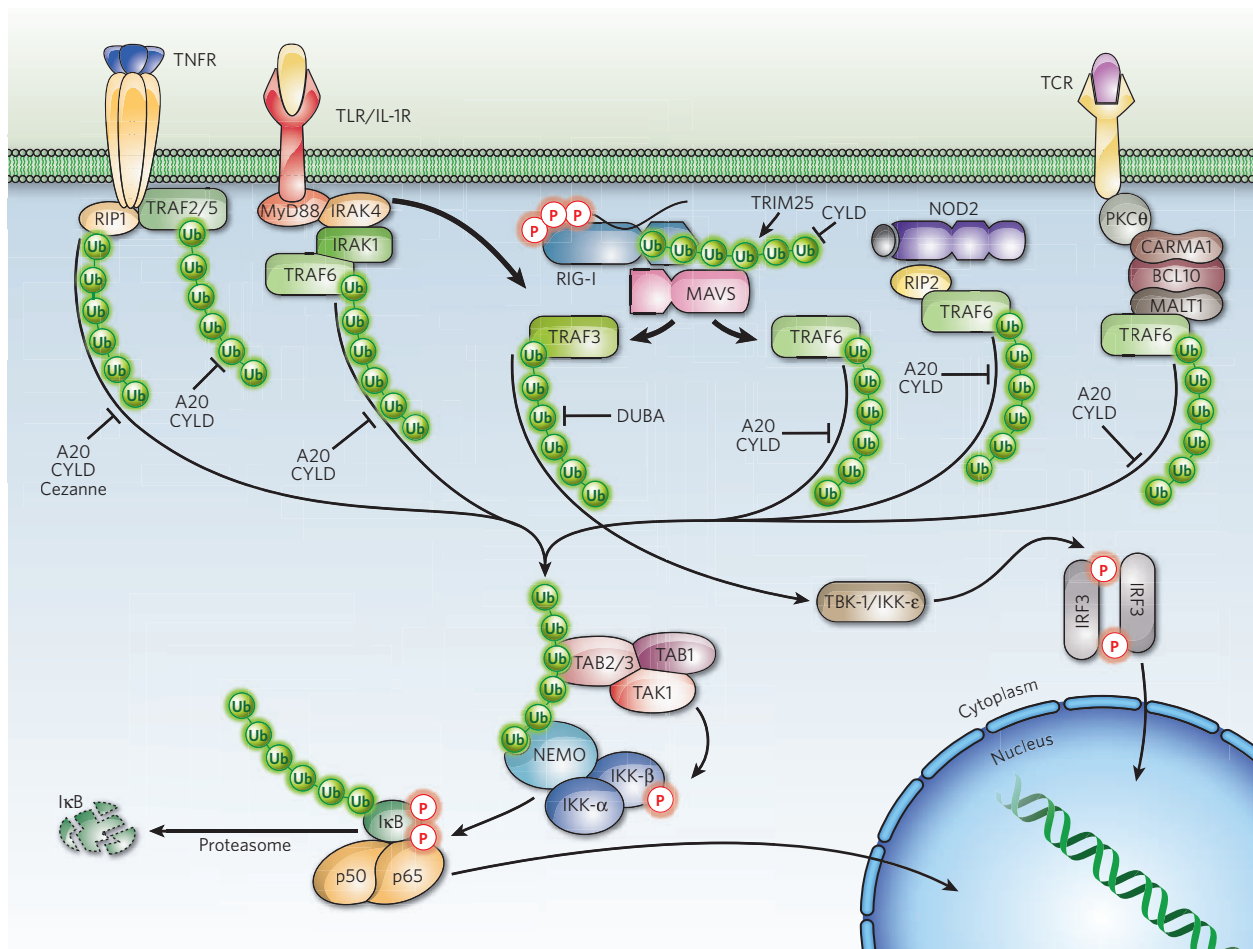
TNF $\alpha$ -induced apoptosis is also inhibited by cIAPs through a distinct ubiquitin-dependent mechanism. Both cIAP1 and cIAP2 bind to SMAC (also known as DIABLO), a mitochondrial protein released to the cytosol under conditions that trigger apoptosis<sup>27</sup>. The binding of SMAC, or small-molecule SMAC mimetics, to cIAPs induces the dimerization, autoubiquitylation and proteasomal degradation of these ubiquitin ligases<sup>28–30</sup>. In the absence of cIAPs, the receptor-associated RIP1 in complex I is rapidly deubiquitylated by CYLD, such that RIP1 is released to the cytosol to form complex II, which initiates apoptosis<sup>31</sup>. cIAPs are also ubiquitin ligases that target the kinase NIK for proteasomal degradation<sup>29</sup>. In the absence of cIAPs, NIK is stabilized and activates the non-canonical pathway of NF- $\kappa$ B to turn on the production of TNF- $\alpha$ , which further promotes apoptotic cell death in an autocrine manner.

### TLR signalling

The TLRs, named after the Toll protein in *Drosophila*, consist of 12 members (TLR1–9 and TLR11–13) in mice and 10 in humans (TLR1–10)<sup>32</sup>. They are membrane-bound receptors containing a cytoplasmic Toll–IL-1 receptor (TIR) domain, which mediates interaction with signalling adaptors. TLRs are often subdivided according to their cellular localization. Endosomal TLRs (TLRs 3, 7, 8 and 9) traffic from the endoplasmic reticulum to the endosome, where they may engage luminal PAMPs from endocytosed microbes. The remaining TLRs are situated on the plasma membrane, where they sense extracellular PAMPs. Each TLR recognizes a distinct PAMP, and together they can detect a broad array of pathogens including bacteria, viruses, fungi and protists. Signalling initiated by these receptors culminates in the activation of NF- $\kappa$ B and, in certain cases (TLRs 3, 4, 7, 8 and 9), IRFs.

Signals initiated by TLRs and IL-1R are transduced through their TIR and follow a similar pathway (Fig. 1). The TIR of these receptors, with the exception of TLR3, interacts with the TIR of the adaptor MyD88. MyD88 in turn recruits proteins of the IRAK family. IRAK4 recruited to the receptor complex phosphorylates and activates IRAK1 and IRAK2, which then promote the oligomerization of TRAF6, which activates its ubiquitin ligase activity. As discussed earlier, together with the E2 complex UBC13–UEV1A, TRAF6 catalyses Lys 63 polyubiquitylation, leading to the activation of TAK1, IKK and MAPK.

TLR3 and TLR4, upon binding of their ligands (double-stranded RNA and lipopolysaccharide, respectively), interact with another



**Figure 3 | The expanding role of Lys 63 polyubiquitylation in innate and adaptive immunity.** In addition to the IL-1R/TLR and TNFR pathways described here and in Figs 1 and 2, Lys-63-linked polyubiquitylation has a key role in signalling from other receptors of the immune systems, including those from T-cell receptors (TCRs), NLRs (such as NOD2) and RLRs (such as RIG-I). In each case, TRAF proteins, including TRAF6, function as ubiquitin ligases to catalyse Lys 63 polyubiquitylation, leading to activation of the TAK1

and IKK complexes. In the RIG-I pathway, TRIM25 catalyses the Lys 63 polyubiquitylation of RIG-I, which is important for the activation of MAVS. MAVS activates not only NF- $\kappa$ B, but also IRF3, which leads to the induction of type-I interferons. The activation of IRF3 requires TRAF3 and IKK-like kinases, TBK1 and IKK- $\epsilon$ . The Lys63 polyubiquitin chains formed in each pathway can be disassembled by DUBs such as A20, Cezanne, CYLD and DUBA, providing a mechanism to downregulate immune responses.

TIR-containing adaptor called TRIF, which binds to TRAF6 and RIP1 to activate the NF- $\kappa$ B pathway. TRIF also binds to the IKK-like kinases TBK1 and IKK- $\epsilon$ , which phosphorylate IRF3. Phosphorylated IRF3 dimerizes and enters the nucleus where it associates with NF- $\kappa$ B and activating transcription factor 2 (ATF2) to form an enhanceosome complex that induces IFN- $\beta$ . Other TLRs that can induce IFN-I include TLR7, TLR8 (human only) and TLR9. TLR7 and TLR8 bind to single-stranded RNA derived from viruses, whereas TLR9 binds unmethylated CpG DNA. This binding recruits a cytosolic signalling complex consisting of MyD88, IRAKs and TRAF6. This complex, when assembled on the endosomal membranes of specialized dendritic cells, such as plasmacytoid dendritic cells (pDCs), activates not only NF- $\kappa$ B, but also IRF7, which is a master regulator of IFN-I, especially IFN- $\alpha$ . The activation of IRF7 requires UBC13 and an intact RING-domain of TRAF6, suggesting that polyubiquitylation is also important for the activation of the IRF family of transcription factors<sup>33</sup>. However, the mechanisms by which ubiquitylation mediates IRF activation by the endosomal TLRs require further investigation.

### NLR signalling

NLRs are a large family of cytosolic proteins that largely conform to a tripartite structure<sup>34</sup>. All NLRs contain a central nucleotide-binding domain (NBD, also known as a NACHT domain) that binds ATP and mediates oligomerization, and a C-terminal LRR region involved in sensing PAMPs or participating in autoinhibition. At their N termini, NLRs have a caspase activation and recruitment domain (CARD), a pyrin domain (PYD), a baculovirus IAP repeat (BIR) or an acidic activation domain; this region is involved in the recruitment of downstream signalling proteins. Two of the best-studied NLRs, NOD1 and NOD2, are found in gut epithelial cells, where they detect the presence of peptidoglycans derived from bacterial cell walls. Mutations in NOD2 have been linked to Crohn's disease, an inflammatory bowel disorder.

Evidence that ubiquitylation is involved in signal transduction by NLRs has been provided by research on NOD1 and NOD2, which can activate NF- $\kappa$ B and MAPK (Fig. 3). Bacterial infection presumably induces the homo-oligomerization of the NOD proteins, allowing them to recruit RIP2 (also known as RICK), a CARD-domain-containing protein kinase. Like RIP1, RIP2 activates IKK through a mechanism that depends on TAK1 and Lys 63 polyubiquitylation<sup>35,36</sup>. The targets of Lys 63 polyubiquitylation are RIP2 and NEMO, and the ubiquitylation of both requires UBC13 and a member of the TRAF family (TRAF2/5 or TRAF6)<sup>35,37–39</sup>. Furthermore, ubiquitylation of RIP2 is restricted by A20, and macrophages lacking A20 exhibit enhanced NF- $\kappa$ B signalling and increased production of proinflammatory cytokines in response to a NOD2 ligand<sup>40</sup>. A Crohn's disease-associated mutation in NOD2 impairs its ability to stimulate NEMO polyubiquitylation as well as its ability to activate IKK<sup>35</sup>. Collectively, these results provide strong evidence that Lys 63 polyubiquitylation has a key role in NOD signalling.

### RLR signalling

There are three RLRs, RIG-I, MDA5 and LGP2, all of which contain RNA helicase domains that are involved in the sensing of viral double-stranded RNA in the cytosol<sup>32</sup>. In addition, RIG-I and LGP2 contain a C-terminal regulatory domain that binds to uncapped 5'-triphosphorylated single-stranded viral RNA, which is distinguished from other cellular RNA that is normally modified or capped at the 5' end. The N termini of RIG-I and MDA5 also contain tandem CARD domains, which interact with the CARD domain of a downstream signalling adaptor MAVS (also known as IPS-1, VISA or CARDIF). LGP2, however, lacks N-terminal CARDs and is thought to have a role in regulation of RIG-I and MDA5 signalling. MAVS contains a transmembrane domain that targets it to the mitochondrial outer membrane, and this localization is important for the activation of IKK and TBK1/IKK- $\epsilon$  in the cytosol. The activation of these kinases by MAVS leads to the activation of NF- $\kappa$ B and IRF3, resulting in the production of IFN-I and other antiviral molecules.

Recent studies indicate that Lys 63 polyubiquitylation has a key role in RIG-I signalling (Fig. 3). Following virus infection, RIG-I is modi-

fied by Lys-63-linked polyubiquitin chains at a lysine residue within the second CARD domain, and this modification enhances the binding of RIG-I to MAVS<sup>41</sup>. The ubiquitylation of RIG-I is catalysed by TRIM25, a RING-domain ubiquitin ligase. Cells lacking TRIM25 are defective in producing type-I interferons in response to viral infection. Additionally, a point mutation of the RIG-I ubiquitylation site impairs its ability to induce IFN-I. Further support for the important role of ubiquitylation in RIG-I activation comes from the finding that CYLD can inhibit the viral induction of IFN-I by reversing the Lys-63-linked polyubiquitylation of RIG-I (refs 42, 43).

Downstream of RIG-I, ubiquitin is also involved in signalling by MAVS. MAVS contains two consensus TRAF6-binding sites and another sequence motif for binding to TRAF2 or TRAF3. The recruitment of TRAF2 and TRAF6 to MAVS is thought to be important for NF- $\kappa$ B activation, whereas the recruitment of TRAF3 leads to the activation of IRF3 (ref. 44). Like TRAF2 and TRAF6, TRAF3 contains an N-terminal RING-domain that enables it to activate IRF3 (ref. 44). Consistent with TRAF3 polyubiquitylation having an important role in MAVS signalling, DUBA, a newly identified DUB, specifically disassembles the Lys-63-linked polyubiquitin chains on TRAF3, resulting in the inhibition of TBK1 and IRF3 (ref. 45).

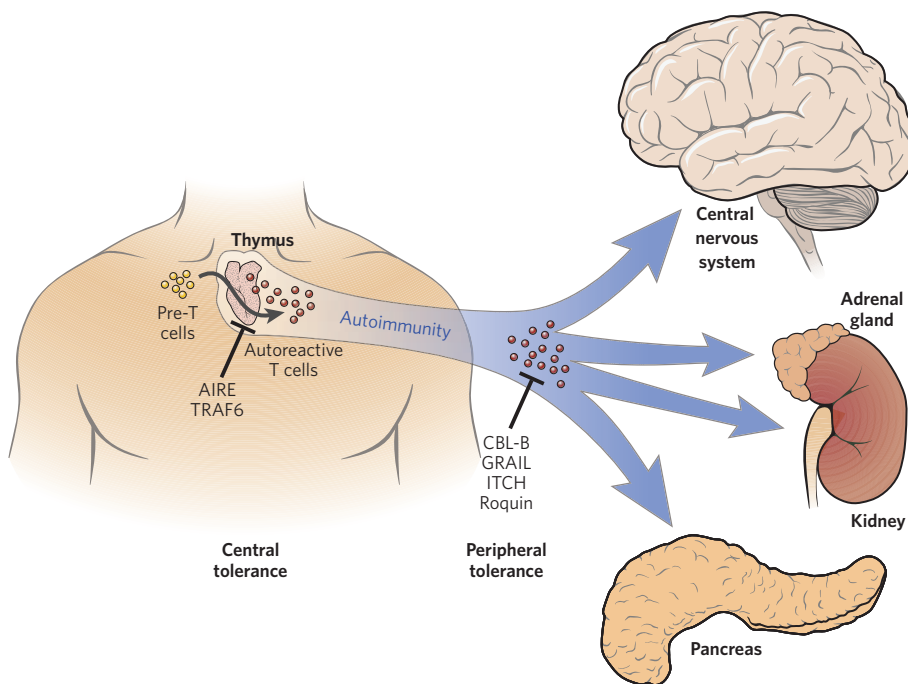
### Ubiquitin in adaptive immunity

Adaptive immunity is under the control of T and B cells, which express an extremely wide range of antigen receptors. T-cell receptors (TCRs) recognize foreign peptide-major histocompatibility complex (MHC) complexes presented by antigen-presenting cells (APCs), which include dendritic cells. When activated by PAMPs, dendritic cells present pathogen peptide-bound MHC as well as other T-cell co-stimulatory molecules on their surface. The initial activation, or priming, of T cells requires MHC peptide-TCR interactions along with interactions between co-stimulatory ligands and their receptors. The priming of T cells initiates the adaptive phase of an immune response. T cells that receive adequate MHC and co-stimulation clonally expand and carry out effector functions. CD8<sup>+</sup> T cells have an important role in antiviral defence by engaging infected cells through interactions between MHC-I and TCRs and lysing them directly. CD4<sup>+</sup> T cells function primarily as helper cells by secreting cytokines to direct the immune response mediated by other effector cells, such as CD8<sup>+</sup> T cells and B cells. B cells, which form the other arm of the adaptive response, are activated by direct contact with antigen through their B-cell receptors. Upon activation, B cells proliferate, mature into plasma cells and secrete antibodies.

### TCR signalling

When engaged by an MHC bound to a cognate peptide epitope, the TCR complex initiates a cascade of tyrosine phosphorylation, resulting in the activation of the serine-threonine kinase PKC $\theta$ . PKC $\theta$  induces the formation of a complex known as CBM, consisting of CARMA1, BCL10 and MALT1 (Fig. 3). The formation of this complex eventually leads to the activation of IKK. MALT1 contains binding sites for both TRAF2 and TRAF6, and recruitment of TRAF6 to the CBM complex induces the E3 ligase activity of TRAF6 by promoting its oligomerization<sup>46</sup>. Several ubiquitylation targets of TRAF6 have been reported, including BCL10, MALT1, NEMO and TRAF6 itself<sup>46–49</sup>. Lys 63 polyubiquitylation of these proteins has been proposed as a mechanism for recruiting and activating the TAK1 and IKK complexes in response to TCR stimulation. The roles of TAK1 and Lys 63 polyubiquitylation in this TCR signalling pathway have been demonstrated using mouse models lacking TAK1 or UBC13 in T-cell lineages<sup>50–53</sup>. Genetic ablation of TAK1 or UBC13 leads to a severe defect in the activation of IKK and MAPK by TCR stimulation in thymocytes and a significant loss of peripheral T cells. However, one study has shown that conditional deletion of UBC13 in other cell types, such as macrophages and fibroblasts, inhibits MAPK but not NF- $\kappa$ B activation by IL-1 $\beta$  and TNF- $\alpha$ <sup>54</sup>, whereas another study showed that the deletion of one allele of UBC13 inhibits NF- $\kappa$ B activation by multiple proinflammatory stimuli<sup>55</sup>. Thus,





**Figure 4 | Ubiquitin ligases in the prevention of autoimmunity.** Pre-T cells developing in the thymus are screened so that most potentially auto-reactive cells are deleted and do not reach the periphery. Thymic screening, also termed 'central tolerance', is orchestrated primarily by medullary thymic epithelial cells, which require the E3 ligases AIRE and TRAF6 for their proper function. Autoreactive T cells that escape the thymus are restricted by various mechanisms of T-cell both intrinsic and extrinsic 'peripheral tolerance'. These mechanisms are maintained by the E3 ligases, CBL-B, GRAIL, ITCH and Roquin. Mutations in any of these centrally or peripherally acting ubiquitin ligases result in the activation of autoreactive T cells, which can mediate destruction of various tissues in an autoimmune response.

the requirement of UBC13 in the NF- $\kappa$ B pathways may depend on cell types and stimuli. It is possible that in some pathways another E2, such as UBC5, may replace UBC13 to mediate NF- $\kappa$ B activation<sup>6</sup>.

### Ubiquitin in autoimmunity

T-cell precursors originating in the bone marrow migrate to the thymus where they undergo an intricate selection process before becoming mature T cells that seed the body. Thymic T-cell selection aims to restrict the maturation of only those cells that will be activated when their TCRs encounter foreign peptides in the context of the MHC. Most self-peptide-MHC-recognizing T cells are deleted to prevent autoimmune activation. This latter function, termed 'central tolerance' is imperfect, allowing the maturation of some potentially autoreactive T cells. These T cells do not normally pose a threat, thanks to an additional system of checks termed 'peripheral tolerance', which is maintained through antigen-induced cell death or the inactivation (anergy) of the T cell in the absence of co-stimulation. Autoreactive T cells may also be suppressed by the action of regulatory T ( $T_{reg}$ ) cells. Several E3 ubiquitin ligases have been implicated in the maintenance of central and peripheral T-cell tolerance (Fig. 4).

### Ubiquitin ligases in central tolerance

In the thymus, T cells that have productively generated a TCR that recognizes the host's MHC survive 'positive selection'. Then, to ensure that dangerous autoreactive T cells do not enter the periphery where they might target self tissues, the thymus deletes those cells that bind to self-peptide-MHC complexes with high affinity in a process known as 'negative selection'. Thymic dendritic cells and medullary epithelial cells (mTEC) present a broad array of self peptides on their MHC for the negative selection of T cells. Mature T cells leaving the thymus are tolerized to the array of peripheral tissue antigens (PTAs). The expression of PTAs in mTECs is controlled by the nuclear protein AIRE, mutations in which have been linked to the human autoimmune disease APECED (also known as APS1)<sup>56,57</sup>. AIRE contains a putative N-terminal CARD, followed by a SAND domain and two PHD motifs. Recombinant full-length AIRE or the PHD1 alone can, in conjunction with an E2 (UBC4 or UBC5b), induce ubiquitylation *in vitro*. Mutations in PHD1 found in patients with APECED have been shown to abolish the E3 activity<sup>58</sup>. However, this result has been disputed<sup>59</sup>, so whether AIRE functions as a physiologically relevant E3 remains to be clarified.

As with AIRE-deficient mice, mice lacking TRAF6 develop inflammatory infiltrates in various organs<sup>60</sup>. Deletion of TRAF6 disrupts thymic organization and mTEC maturation, and results in reduced thymic AIRE expression. TRAF6 and NIK are required for REL-B expression, which is in turn required for thymic AIRE expression. Recent studies have shown that CD40 and RANK mediate the activation of TRAF6, NIK and IKK- $\beta$  in mTECs<sup>61</sup>.

### Ubiquitin ligases in peripheral T-cell tolerance

Cbl-b is a member of the Casitas B-lineage lymphoma family of proteins and is a RING-containing E3 known to regulate proximal TCR signaling<sup>62,63</sup>. Disruption of the gene that encodes Cbl-b in mice leads to hyperactivation of PCK0, Akt and NF- $\kappa$ B. T cells that lack Cbl-b can be activated independently of CD28 co-stimulation and so are resistant to the induction of anergy. Of the Cbl-b ubiquitylation targets, phospholipase C- $\gamma$ 1 (PLC- $\gamma$ 1) seems most relevant to anergy induction<sup>64,65</sup>. During anergy induction of T cells, Cbl-b ubiquitylates PLC- $\gamma$ 1, resulting in its inactivation or degradation, thereby preventing the T cells from being fully activated. Cbl-b is targeted for degradation by the HECT-domain ubiquitin ligase Nedd4, which promotes T-cell activation<sup>66</sup>.

GRAIL is another RING-domain E3 that has a role in peripheral tolerance<sup>67</sup>. When stimulated, T-cell hybridomas that ectopically express GRAIL show decreased IL-2 and IL-4 transcript and protein levels. This anergic response is dependent on GRAIL's RING domain, suggesting the importance of its ligase activity. Increased GRAIL levels have been found in anergized CD4<sup>+</sup> T cells, as well as in CD4<sup>+</sup>CD25<sup>+</sup>  $T_{reg}$  cells<sup>64</sup>. GRAIL has recently been shown to catalyse Lys-63-linked polyubiquitylation of RhoGDI, which leads to the inhibition of Rho A GTPase activity<sup>68</sup>. GRAIL also promotes the ubiquitylation and degradation of CD40L, which is known to contribute to co-stimulation of T cells<sup>69</sup>.

The HECT-domain E3 ITCH has a key role in peripheral T-cell tolerance<sup>70</sup>. Disruption of ITCH *in vivo* causes a break in T-cell tolerance, resulting in lymphoid hyperplasia. Under tolerizing conditions, T cells deficient in ITCH produce higher levels of  $T_H2$  cytokines, such as IL-4, and show greater proliferation than wild-type cells<sup>71</sup>. ITCH is phosphorylated by JNK, and this phosphorylation enhances its ubiquitin ligase activity, leading to polyubiquitylation and degradation of JunB<sup>72</sup>. In  $T_H2$  cells, JunB binds to the IL-4 gene promoter and is required for its transcription. Therefore, in the absence of ITCH, JunB is stabilized, leading to a skewed  $Th2$  response and autoimmunity. Another

substrate of ITCH is TIEG1, a transcription factor required for TGF $\beta$ -induced expression of FoxP3, which is required for the maintenance of T<sub>reg</sub> cells<sup>73</sup>. ITCH catalyses polyubiquitylation of TIEG1; however, this ubiquitylation activates TIEG1 instead of targeting it for degradation. Thus, ITCH mediates peripheral T-cell tolerance through the induction of T<sub>reg</sub> cells.

Roquin, a putative RING-domain E3, may also promote peripheral tolerance<sup>74</sup>. Roquin-mutant T cells express high levels of IL-21 and surface ICOS, a co-stimulatory molecule. Enhanced ICOS expression leads to a break in tolerance, resulting in a phenotype resembling systemic lupus erythematosus. Roquin contains an N-terminal RING domain followed by a ROQ domain and then a zinc-finger domain similar to those found in proteins that bind to and regulate messenger RNA stability. Roquin may regulate the stability of ICOS mRNA, although it is not known whether this regulation requires Roquin's RING domain.

## Perspectives

The examples presented here provide only a glimpse into the expanding role of ubiquitylation in the regulation of immune responses. It is clear that polyubiquitylation is important not only for terminating immune responses through the destruction of immune regulatory proteins, but also for initiating and amplifying immune responses, as exemplified by the role of Lys-63-linked polyubiquitylation in the activation of protein kinases in the NF- $\kappa$ B and IRF pathways. As more components of the ubiquitin system, such as E3s and DUBs, are linked to the immune system, one challenge will be to understand the biochemical mechanisms underlying immune regulation by ubiquitin. Key to this mechanistic understanding is the identification of the substrates of ubiquitylation. Because de-ubiquitylation and degradation occur rapidly, the steady-state levels of the ubiquitylated intermediates in cells are normally very low. Fortunately, technological advances such as the use of mass spectrometry, which can identify proteins and map modification sites with very high sensitivity, should aid the discovery of the physiological substrates of ubiquitylation.

Equally important to substrate identification is the identification of the 'readers' of the ubiquitin signal. The discovery of nearly 20 types of UBD in the past few years should provide a boost to the identification of ubiquitin receptors involved in the immune system. If we know both the ubiquitylation targets and the ubiquitin receptors, it may be possible to biochemically reconstitute and reconstruct the ubiquitin signalling events as they happen during immune responses *in vivo*. A deep mechanistic understanding of ubiquitin signalling would be extremely valuable for the future development of treatments for immune and inflammatory diseases.

- Chen, Z. J. Ubiquitin signalling in the NF- $\kappa$ B pathway. *Nature Cell Biol.* **7**, 758–765 (2005).
- Chen, M. & Gerlier, D. Viral hijacking of cellular ubiquitination pathways as an anti-innate immunity strategy. *Viral Immunol.* **19**, 349–362 (2006).
- Loureiro, J. & Ploegh, H. L. Antigen presentation and the ubiquitin-proteasome system in host-pathogen interactions. *Adv. Immunol.* **92**, 225–305 (2006).
- Pomerantz, J. L. & Baltimore, D. Two pathways to NF- $\kappa$ B. *Mol. Cell* **10**, 693–695 (2002).
- Palombella, V. J., Rando, O. J., Goldberg, A. L. & Maniatis, T. The ubiquitin-proteasome pathway is required for processing the NF- $\kappa$ B1 precursor protein and the activation of NF- $\kappa$ B. *Cell* **78**, 773–785 (1994).
- Chen, Z. J., Parent, L. & Maniatis, T. Site-specific phosphorylation of I $\kappa$ B $\alpha$  by a novel ubiquitination-dependent protein kinase activity. *Cell* **84**, 853–862 (1996). This study uncovers the I $\kappa$ B kinase complex and shows that kinase activity could be activated by polyubiquitylation *in vitro* through a proteasome-independent mechanism.
- Deng, L. *et al.* Activation of the I $\kappa$ B kinase complex by TRAF6 requires a dimeric ubiquitin-conjugating enzyme complex and a unique polyubiquitin chain. *Cell* **103**, 351–361 (2000). This paper shows that TRAF6 is a RING-domain ubiquitin E3 ligase that functions with Ubc13/Uev1A to catalyse Lys-63-linked polyubiquitylation, leading to IKK activation.
- Wang, C. *et al.* TAK1 is a ubiquitin-dependent kinase of MKK and IKK. *Nature* **412**, 346–351 (2001).
- Kanayama, A. *et al.* TAB2 and TAB3 activate the NF- $\kappa$ B pathway through binding to polyubiquitin chains. *Mol. Cell* **15**, 535–548 (2004).
- Ea, C. K., Deng, L., Xia, Z. P., Pineda, G. & Chen, Z. J. Activation of IKK by TNF $\alpha$  requires site-specific ubiquitination of RIP1 and polyubiquitin binding by NEMO. *Mol. Cell* **22**, 245–257 (2006).
- Wu, C. J., Conze, D. B., Li, T., Srinivasula, S. M. & Ashwell, J. D. Sensing of Lys 63-linked polyubiquitination by NEMO is a key event in NF- $\kappa$ B activation. *Nature Cell Biol.* **8**, 398–406 (2006). This paper and reference 10 show that NEMO contains a ubiquitin-binding domain required for IKK activation.
- Krappmann, D. & Scheidereit, C. A pervasive role of ubiquitin conjugation in activation and termination of I $\kappa$ B kinase pathways. *EMBO Rep.* **6**, 321–326 (2005).
- Sun, S. C. Deubiquitylation and regulation of the immune response. *Nature Rev. Immunol.* **8**, 501–511 (2008).
- Bignell, G. R. *et al.* Identification of the familial cylindromatosis tumour-suppressor gene. *Nature Genet.* **25**, 160–165 (2000).
- Krikos, A., Laherty, C. D. & Dixit, V. M. Transcriptional activation of the tumor necrosis factor  $\alpha$ -inducible zinc finger protein, A20, is mediated by  $\kappa$ B elements. *J. Biol. Chem.* **267**, 17971–17976 (1992).
- Lee, E. G. *et al.* Failure to regulate TNF-induced NF- $\kappa$ B and cell death responses in A20-deficient mice. *Science* **289**, 2350–2354 (2000).
- Wertz, I. E. *et al.* De-ubiquitination and ubiquitin ligase domains of A20 downregulate NF- $\kappa$ B signalling. *Nature* **430**, 694–699 (2004). This paper shows that A20 inhibits IKK through dual enzymatic activities: a DUB activity removes Lys-63-linked polyubiquitin chains from RIP1, then an E3 activity conjugates Lys-48-linked polyubiquitin chains to RIP1, targeting RIP1 for proteasomal degradation.
- Komander, D. & Barford, D. Structure of the A20 OTU domain and mechanistic insights into deubiquitination. *Biochem. J.* **409**, 77–85 (2008).
- Coornaert, B., Carpentier, I. & Beyaert, R. A20: central gatekeeper in inflammation and immunity. *J. Biol. Chem.* doi:10.1074/jbc.R800032200 (published online 13 November 2008).
- Shembade, N. *et al.* The E3 ligase Itch negatively regulates inflammatory signaling pathways by controlling the function of the ubiquitin-editing enzyme A20. *Nature Immunol.* **9**, 254–262 (2008).
- Janeway, C. A. Jr & Medzhitov, R. Innate immune recognition. *Annu. Rev. Immunol.* **20**, 197–216 (2002).
- Micheau, O. & Tschopp, J. Induction of TNF receptor I-mediated apoptosis via two sequential signaling complexes. *Cell* **114**, 181–190 (2003).
- Li, H., Kobayashi, M., Blonska, M., You, Y. & Lin, X. Ubiquitination of RIP is required for tumor necrosis factor  $\alpha$ -induced NF- $\kappa$ B activation. *J. Biol. Chem.* **281**, 13636–13643 (2006).
- Lee, T. H., Shank, J., Cusson, N. & Kelliher, M. A. The kinase activity of Rip1 is not required for tumor necrosis factor- $\alpha$ -induced I $\kappa$ B kinase or p38 MAP kinase activation or for the ubiquitination of Rip1 by Traf2. *J. Biol. Chem.* **279**, 33185–33191 (2004).
- Bertrand, M. J. *et al.* cIAP1 and cIAP2 facilitate cancer cell survival by functioning as E3 ligases that promote RIP1 ubiquitination. *Mol. Cell* **30**, 689–700 (2008).
- Chang, L. *et al.* The E3 ubiquitin ligase Itch couples JNK activation to TNF $\alpha$ -induced cell death by inducing c-FLIP<sub>L</sub> turnover. *Cell* **124**, 601–613 (2006).
- Du, C., Fang, M., Li, Y., Li, L. & Wang, X. Smac, a mitochondrial protein that promotes cytochrome c-dependent caspase activation by eliminating IAP inhibition. *Cell* **102**, 33–42 (2000).
- Petersen, S. L. *et al.* Autocrine TNF $\alpha$  signaling renders human cancer cells susceptible to Smac-mimetic-induced apoptosis. *Cancer Cell* **12**, 445–456 (2007).
- Varfolomeev, E. *et al.* IAP antagonists induce autoubiquitination of c-IAPs, NF- $\kappa$ B activation, and TNF $\alpha$ -dependent apoptosis. *Mol. Cell* **131**, 669–681 (2007).
- Vince, J. E. *et al.* IAP antagonists target cIAP1 to induce TNF $\alpha$ -dependent apoptosis. *Cell* **131**, 682–693 (2007).
- Wang, L., Du, F. & Wang, X. TNF- $\alpha$  induces two distinct caspase-8 activation pathways. *Cell* **133**, 693–703 (2008).
- Pichlmair, A. & Reis e Sousa, C. Innate recognition of viruses. *Immunity* **27**, 370–383 (2007).
- Kawai, T. *et al.* Interferon- $\alpha$  induction through Toll-like receptors involves a direct interaction of IRF7 with MyD88 and TRAF6. *Nature Immunol.* **5**, 1061–1068 (2004).
- Kanneganti, T. D., Lamkanfi, M. & Nunez, G. Intracellular NOD-like receptors in host defense and disease. *Immunity* **27**, 549–559 (2007).
- Abbott, D. W., Wilkins, A., Asara, J. M. & Cantley, L. C. The Crohn's disease protein, NOD2, requires RIP2 in order to induce ubiquitylation of a novel site on NEMO. *Curr. Biol.* **14**, 2217–2227 (2004).
- Windheim, M., Lang, C., Pegg, M., Plater, L. A. & Cohen, P. Molecular mechanisms involved in the regulation of cytokine production by muramyl dipeptide. *Biochem. J.* **404**, 179–190 (2007).
- Abbott, D. W. *et al.* Coordinated regulation of Toll-like receptor and NOD2 signaling by K63-linked polyubiquitin chains. *Mol. Cell Biol.* **27**, 6012–6025 (2007).
- Hasegawa, M. *et al.* A critical role of RICK/RIP2 polyubiquitination in Nod-induced NF- $\kappa$ B activation. *EMBO J.* **27**, 373–383 (2008).
- Yang, Y. *et al.* NOD2 pathway activation by MDP or *Mycobacterium tuberculosis* infection involves the stable polyubiquitination of Rip2. *J. Biol. Chem.* **282**, 36223–36229 (2007).
- Hitotsumatsu, O. *et al.* The ubiquitin-editing enzyme A20 restricts nucleotide-binding oligomerization domain containing 2-triggered signals. *Immunity* **28**, 381–390 (2008).
- Gack, M. U. *et al.* TRIM25 RING-finger E3 ubiquitin ligase is essential for RIG-I-mediated antiviral activity. *Nature* **446**, 916–920 (2007). This paper shows that Lys-63-linked polyubiquitylation of RIG-I by TRIM25 has an important role in triggering type-I interferon production in response to viral infection.
- Friedman, C. S. *et al.* The tumour suppressor CYLD is a negative regulator of RIG-I-mediated antiviral response. *EMBO Rep.* **9**, 930–936 (2008).
- Zhang, M. *et al.* Regulation of I $\kappa$ B kinase-related kinases and antiviral responses by tumor suppressor CYLD. *J. Biol. Chem.* **283**, 18621–18626 (2008).
- Saha, S. K. *et al.* Regulation of antiviral responses by a direct and specific interaction between TRAF3 and Cardif. *EMBO J.* **25**, 3257–3263 (2006).
- Kayagaki, N. *et al.* DUBA: a deubiquitinase that regulates type I interferon production. *Science* **318**, 1628–1632 (2007).
- Sun, L., Deng, L., Ea, C. K., Xia, Z. P. & Chen, Z. J. The TRAF6 ubiquitin ligase and TAK1 kinase mediate IKK activation by BCL10 and MALT1 in T lymphocytes. *Mol. Cell* **14**, 289–301 (2004).
- Oeckinghaus, A. *et al.* Malt1 ubiquitination triggers NF- $\kappa$ B signaling upon T-cell activation. *EMBO J.* **26**, 4634–4645 (2007).
- Wu, C. J. & Ashwell, J. D. NEMO recognition of ubiquitinated Bcl10 is required for T cell receptor-mediated NF- $\kappa$ B activation. *Proc. Natl Acad. Sci. USA* **105**, 3023–3028 (2008).



49. Zhou, H. *et al.* Bcl10 activates the NF- $\kappa$ B pathway through ubiquitination of NEMO. *Nature* **427**, 167–171 (2004).  
This paper and reference 46 show that Lys-63-linked polyubiquitylation is important for IKK activation by BCL10 and MALT1 in T cells.
50. Yamamoto, M. *et al.* Pivotal function of Ubc13 in thymocyte TCR signaling. *J. Immunol.* **177**, 7520–7524 (2006).
51. Wan, Y. Y., Chi, H., Xie, M., Schneider, M. D. & Flavell, R. A. The kinase TAK1 integrates antigen and cytokine receptor signaling for T cell development, survival and function. *Nature Immunol.* **7**, 851–858 (2006).
52. Sato, S. *et al.* TAK1 is indispensable for development of T cells and prevention of colitis by the generation of regulatory T cells. *Int. Immunol.* **18**, 1405–1411 (2006).
53. Liu, H. H., Xie, M., Schneider, M. D. & Chen, Z. J. Essential role of TAK1 in thymocyte development and activation. *Proc. Natl Acad. Sci. USA* **103**, 11677–11682 (2006).
54. Yamamoto, M. *et al.* Key function for the Ubc13 E2 ubiquitin-conjugating enzyme in immune receptor signaling. *Nature Immunol.* **7**, 962–970 (2006).
55. Fukushima, T. *et al.* Ubiquitin-conjugating enzyme Ubc13 is a critical component of TNF receptor-associated factor (TRAF)-mediated inflammatory responses. *Proc. Natl Acad. Sci. USA* **104**, 6371–6376 (2007).
56. Aaltonen, J. *et al.* An autoimmune disease, APECED, caused by mutations in a novel gene featuring two PHD-type zinc-finger domains. *Nature Genet.* **17**, 399–403 (1997).
57. Nagamine, K. *et al.* Positional cloning of the APECED gene. *Nature Genet.* **17**, 393–398 (1997).
58. Uchida, D. *et al.* AIRE functions as an E3 ubiquitin ligase. *J. Exp. Med.* **199**, 167–172 (2004).
59. Bottomley, M. J. *et al.* NMR structure of the first PHD finger of autoimmune regulator protein (AIRE1). Insights into autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) disease. *J. Biol. Chem.* **280**, 11505–11512 (2005).
60. Akiyama, T. *et al.* Dependence of self-tolerance on TRAF6-directed development of thymic stroma. *Science* **308**, 248–251 (2005).
61. Akiyama, T. *et al.* The tumor necrosis factor family receptors RANK and CD40 cooperatively establish the thymic medullary microenvironment and self-tolerance. *Immunity* **29**, 423–437 (2008).
62. Chiang, Y. J. *et al.* Cbl-b regulates the CD28 dependence of T-cell activation. *Nature* **403**, 216–220 (2000).
63. Bachmaier, K. *et al.* Negative regulation of lymphocyte activation and autoimmunity by the molecular adaptor Cbl-b. *Nature* **403**, 211–216 (2000).
64. Heissmeyer, V. *et al.* Calcineurin imposes T cell unresponsiveness through targeted proteolysis of signaling proteins. *Nature Immunol.* **5**, 255–265 (2004).
65. Jeon, M. S. *et al.* Essential role of the E3 ubiquitin ligase Cbl-b in T cell anergy induction. *Immunity* **21**, 167–177 (2004).
66. Yang, B. *et al.* Nedd4 augments the adaptive immune response by promoting ubiquitin-mediated degradation of Cbl-b in activated T cells. *Nature Immunol.* **9**, 1356–1363 (2008).
67. Anandasabapathy, N. *et al.* GRAIL: an E3 ubiquitin ligase that inhibits cytokine gene transcription is expressed in anergic CD4<sup>+</sup> T cells. *Immunity* **18**, 535–547 (2003).
68. Su, L., Lineberry, N., Huh, Y., Soares, L. & Fathman, C. G. A novel E3 ubiquitin ligase substrate screen identifies Rho guanine dissociation inhibitor as a substrate of gene related to anergy in lymphocytes. *J. Immunol.* **177**, 7559–7566 (2006).
69. Lineberry, N. B. *et al.* The transmembrane E3 ligase GRAIL ubiquitinates the costimulatory molecule CD40 ligand during the induction of T cell anergy. *J. Immunol.* **181**, 1622–1626 (2008).
70. Liu, Y. C. The E3 ubiquitin ligase Itch in T cell activation, differentiation, and tolerance. *Semin. Immunol.* **19**, 197–205 (2007).
71. Fang, D. *et al.* Dysregulation of T lymphocyte function in itchy mice: a role for Itch in T<sub>H</sub>2 differentiation. *Nature Immunol.* **3**, 281–287 (2002).  
This paper shows that Itch associates with and ubiquitinates JunB, a key transcription factor involved in T<sub>H</sub>2 differentiation.
72. Gao, M. *et al.* Jun turnover is controlled through JNK-dependent phosphorylation of the E3 ligase Itch. *Science* **306**, 271–275 (2004).
73. Venuprasad, K. *et al.* The E3 ubiquitin ligase Itch regulates expression of transcription factor Foxp3 and airway inflammation by enhancing the function of transcription factor TIEG1. *Nature Immunol.* **9**, 245–253 (2008).
74. Vinuesa, C. G. *et al.* A RING-type ubiquitin ligase family member required to repress follicular helper T cells and autoimmunity. *Nature* **435**, 452–458 (2005).

**Acknowledgements** We apologize to the many authors whose publications are not cited directly because of space limitations. We thank J. Cabrera for graphic support. Research in our laboratory is supported by the Howard Hughes Medical Institute and by grants from the US National Institutes of Health and the Welch Foundation.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Correspondence should be addressed to Z.J.C. ([zhijian.chen@utsouthwestern.edu](mailto:zhijian.chen@utsouthwestern.edu)).

# Targeting the ubiquitin system in cancer therapy

Daniela Hoeller<sup>1</sup> & Ivan Dikic<sup>2,3,4</sup>

**The ubiquitin system is a network of proteins dedicated to the ubiquitylation of cellular targets and the subsequent control of numerous cellular functions. The deregulation of components of this elaborate network leads to human pathogenesis, including the development of many types of tumour. Alterations in the ubiquitin system that occur during the initiation and progression of cancer are now being uncovered, and this knowledge is starting to be exploited for both molecular diagnostics and the development of novel strategies to combat cancer.**

Ubiquitin is a small modifier molecule that labels proteins in a highly specific manner. Like phosphorylation, ubiquitylation — the modification of proteins with ubiquitin — is prevalent in normal, as well as pathological, cellular processes. Initially, ubiquitylation was described as the process that labels proteins for degradation by the proteasome<sup>1</sup>. With time, however, an increasing number of distinct functions have been assigned to different types of ubiquitin modification (for example, the addition of a single ubiquitin molecule or a ubiquitin chain), such as the regulation of protein trafficking, the assembly of protein signalling complexes, and the activation or inactivation of enzymes<sup>2–5</sup>.

The attachment of ubiquitin to a substrate requires the consecutive action of three enzymes (reviewed in ref. 6). The first step involves the activation of ubiquitin by the formation of a thioester bond with the ubiquitin-activating enzyme, E1. In the second step, E1 delivers the activated ubiquitin to the E2 ubiquitin-conjugating enzyme. Finally, E3 ligases catalyse the transfer of ubiquitin from E2 to a lysine residue in the substrate protein. Ubiquitin contains seven acceptor lysines that can participate in ubiquitin conjugation, giving rise to ubiquitin chains of different topologies, lengths and functional outcomes<sup>7</sup>. For example, ubiquitin chains with conjugation through the lysine at position 48 of ubiquitin (Lys 48) lead to proteasomal degradation of the modified substrate, whereas chains linked at Lys 63 are instead implicated in signalling or trafficking events<sup>2</sup>.

Many proteins regulated by ubiquitylation control cellular processes relevant to tumorigenesis, such as cell-cycle progression, apoptosis, receptor downregulation and gene transcription. In recent years, substantial progress has been made in understanding the molecular basis of ubiquitin action in cancer-relevant processes. Here we review the components of the ubiquitin system associated with cancer and discuss their potential as targets for therapeutic strategies against the disease (Fig. 1).

## The ubiquitin ligase system is a target of oncogenic signals

The E3 ligases are considered to be the most important components of the ubiquitin conjugation machinery, as they bind directly to their target proteins and so have substrate specificity. Ubiquitin ligases fall into different classes, based on their structural composition and mechanism of action, and are characterized by the presence of either a HECT, U-box or RING domain. In some cases, the core of the ligase is composed of several proteins, for example in the SCF complex, where SKP1, CUL1,

RBX1 and a variable, substrate-specific F-box protein together form a multi-subunit ubiquitin ligase<sup>6</sup>. Several E3 ligases have been linked to the development of cancer (reviewed in refs 8–10), largely because of their ability to trigger the degradation of oncogenes or tumour suppressors. Well-studied E3 enzymes with oncogenic capacity are the RING-type E3 enzyme Hdm2, a crucial negative regulator of the tumour-suppressor protein p53 (reviewed in refs 11,12), and the multi-subunit SCF ligases that regulate cell-cycle progression. Depending on its substrate specificity, an E3 ligase can act as either a tumour promoter or suppressor, and may function as an oncoprotein in one tissue type despite having tumour-suppressive effects in others.

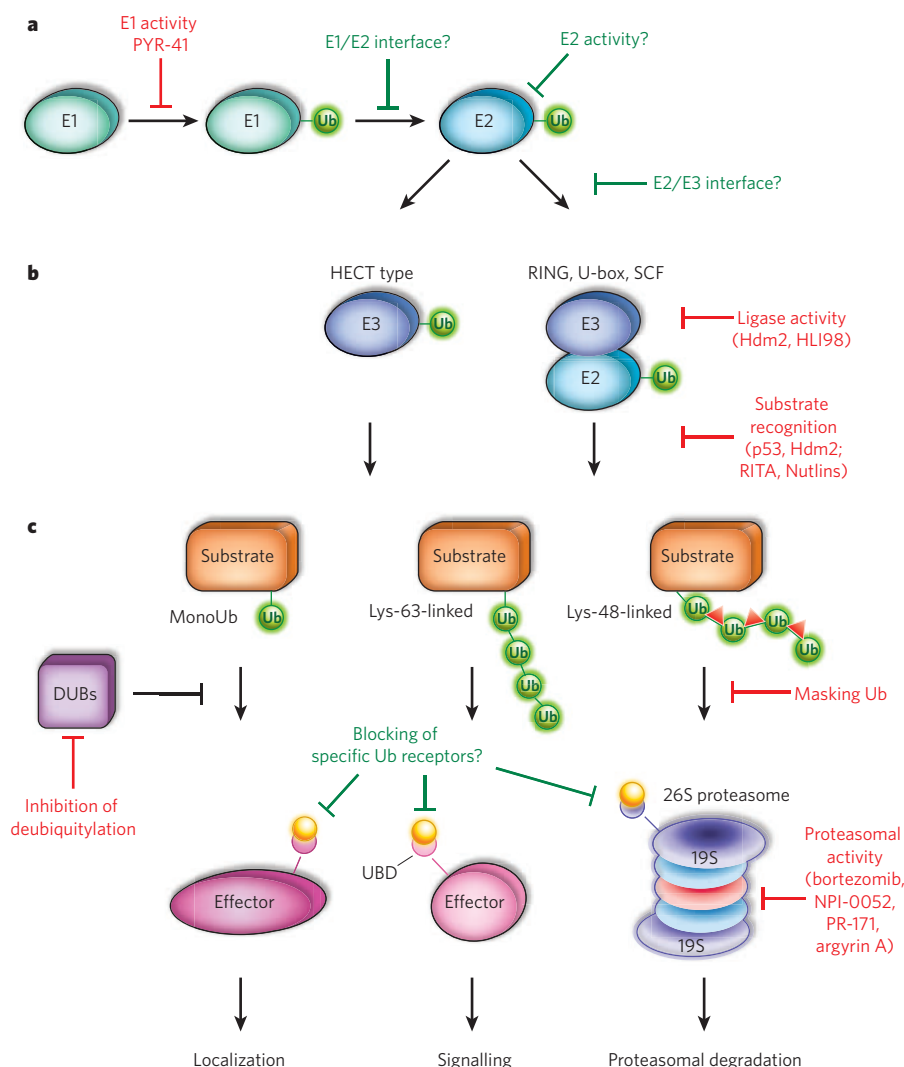
## Targeting ubiquitin ligases

Ubiquitin ligases mediate the degradation of specific substrates, so it was initially believed that targeting the active site of E3 enzymes or their interaction with substrates would create selective drugs with fewer side effects. One obvious target was Hdm2, whose inactivation in tumours should activate the p53 pathway, leading to cell-cycle arrest and apoptosis; another was SKP2, a substrate-specific subunit of the SCF ligase complex whose inactivation in tumours expressing low levels of the cell-cycle inhibitor p27 is likely to be beneficial. On the other hand, there are several reasons to instead promote the activity of E3 enzymes, even though this is more challenging than designing inhibitors. For instance, using agonists of Fbxw7 in tumours expressing high levels of c-Myc or cyclin E should lead to the degradation of these specific oncoproteins, and promoting the activity of VHL ligase to destabilize HIF1 $\alpha$  could block tumour vascularization. In the past decade, the biotech and pharmaceutical industries have sought to develop such inhibitors and agonists of ubiquitin ligases (reviewed in refs 8, 13–16). However, none of these approaches has yielded a significant clinical advance in treating selected tumours.

Why have these evidence-based efforts been unsuccessful, and what can we learn from them to improve future studies? In the case of Hdm2, various approaches have been attempted, including the inhibition of Hdm2 expression, blocking ubiquitin ligase activity, and preventing the interaction between Hdm2 and p53 (Fig. 1b; reviewed in ref. 14). In addition, small-molecule inhibitors have been generated that specifically target the E3 ligase activity of Hdm2 (ref. 17). Indeed, treatment of cancer cells with HLI98 inhibitors has been shown to activate p53 signalling

<sup>1</sup>Division of Medical Biochemistry, Innsbruck Medical University, Biocenter, Fritz-Pregl-Strasse 3, 6020 Innsbruck, Austria. <sup>2</sup>Institute of Biochemistry II and Cluster of Excellence 'Macromolecular Complexes', Goethe University, Theodor-Stern-Kai 7, 60590 Frankfurt, Germany. <sup>3</sup>Tumor Biology Program, Mediterranean Institute for Life Sciences, 21000 Split, Croatia. <sup>4</sup>Department of Immunology, School of Medicine, University of Split, Soltanska 2, 21000 Split, Croatia.





**Figure 1 | The ubiquitin system offers several possibilities for therapeutic intervention.** Known points of intervention and appropriate drugs are shown in red; potential novel points of interference in green. **a**, Global changes can be achieved by blocking early steps of the conjugation machinery, for example at the level of the E1 and E2 enzymes that activate ubiquitin (Ub). **b**, The inhibition of distinct E3-enzyme activities allows for more specific interference. By targeting the E3-substrate interface, alterations can be further confined to a particular substrate (for example, by use of RITA or the Nutlins). This principle also accounts for the actions of deubiquitylating enzymes (DUBs). **c**, Ubiquitin receptors then recognize the ubiquitylated substrate and mediate the cellular response, or DUBs can deubiquitylate the substrate. These actions can be targeted with inhibitors that mask the ubiquitin modification, as is the case with ubistatins (red triangles), which specifically bind Lys-48-linked chains. Another highly specific way of interfering would be to occupy the ubiquitin-binding domains (UBDs) of specific effector proteins. Proteins carrying Lys-48-linked polyubiquitin chains are recognized and degraded by the proteasome. In addition to substances that block enzymatic activity (such as bortezomib, NPI-0052, PR-171 and argyran A), drugs can target the receptors for polyubiquitylated substrates, such as Rpn10 (localized on the proteasome), or shuttling receptors, such as HRad23 (indicated in yellow).

and induce apoptosis in a p53-dependent manner. However, these compounds have low biological potency and considerable p53-independent off-target effects<sup>14</sup>. For instance, several chemical compounds, the most successful being the Nutlins, were identified to target the groove on Hdm2 where p53 loops bind<sup>14</sup>. Although the initial results with Nutlins were promising, more recent studies have uncovered some shortcomings that might preclude further development of this class of inhibitors. Specifically, Nutlins can only be used in tumour cells with wild-type p53, as cells in which p53 is deleted or mutated are insensitive to them. Nutlins have also been shown to induce cell-cycle arrest in a p53-deficient cell line<sup>18</sup>, showing that these inhibitors are not specific for p53 and may compete with other proteins for Mdm2 binding. There is also evidence that other ubiquitin ligases, not inhibited by Nutlins, target p53 for proteasome-dependent degradation (reviewed in ref. 19).

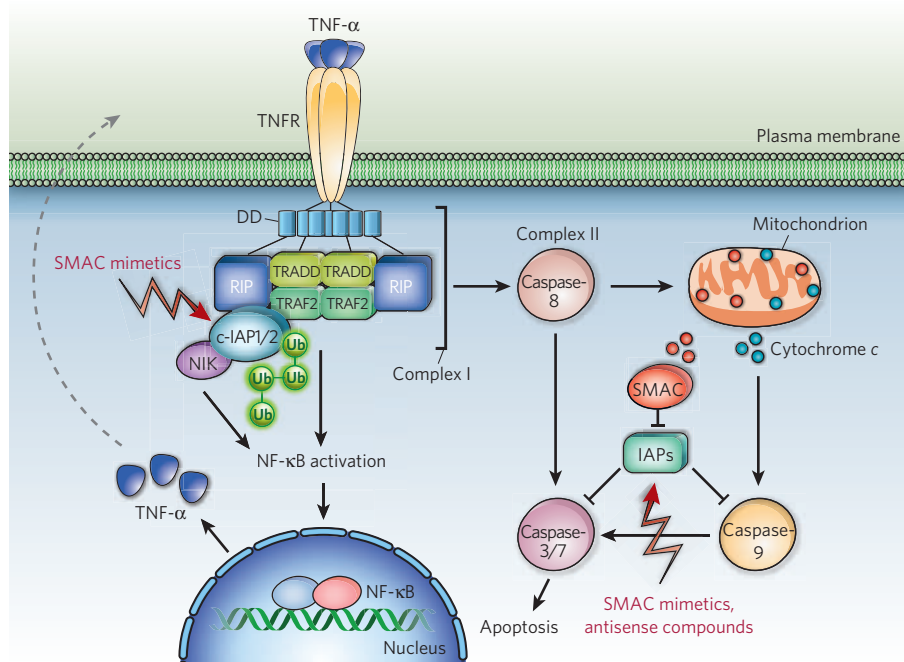
An alternative strategy in the development of ligase inhibitors is to identify small molecules that will bind to the protein substrate and potentially prevent the actions of multiple ligases. Inhibitors could then be created that specifically target wild-type or mutant substrate variants. In the case of p53, small compounds called RITA (2,5-bis(5-hydroxymethyl-2-thienyl)furan) have been shown to bind the amino terminus of p53 and promote growth arrest<sup>20</sup> (Fig. 1b). However, the action of RITA is not specific for p53-Hdm2 interactions; it also affects several other p53 binding partners that inhibit p53 by means different from ubiquitylation (reviewed in ref. 16).

Similar problems in designing peptidomimetics, which could interfere with protein-protein binding surfaces, have been observed for numerous cell-cycle ligases. For example, SCF- $\beta$ -TrCP recognizes two well

defined interspaced phosphorylated serine residues (DS<sub>52</sub>GXXS<sub>56</sub>) in substrates such as  $\beta$ -catenin and I $\kappa$ B $\alpha$ , whereas SCF-Skp2 recognizes a phosphorylated threonine residue in p27<sup>Kip1</sup> (ref. 21). It is still unclear whether interference with such a defined hotspot on a large interaction surface will be sufficient and specific enough for inhibition *in vivo*. Moreover, any small molecule with a phosphate or phosphomimetic is likely to have problems with bioavailability or membrane permeation because of its hydrophilic property. Even if the peptidomimetic compound can be delivered into cells and can inhibit the binding of the ligase to a particular substrate, one needs to consider that most ligases target multiple proteins. For example, interference with  $\beta$ -TrCP can prevent degradation of I $\kappa$ B in tumour cells (which should decrease the pro-oncogenic nuclear factor- $\kappa$ B (NF- $\kappa$ B) pathway), but it may also inhibit the degradation of  $\beta$ -catenin and instead promote cancer or other effects characterized by increased  $\beta$ -catenin levels in cells.

### E3 ligases and tumour vascularization

Tumour vascularization is a hallmark of cancer. It is essential for the rapid expansion of a tumour mass because it secures the supply of nutrients and oxygen and is therefore a rate-limiting secondary step in tumorigenesis<sup>22</sup>. The Elongin B/C-CUL2-VHL complex is a prototypical example of an E3 ligase involved in the regulation of vascularization. The von Hippel-Lindau (VHL) protein is the substrate-specific component in the complex. Mutations in the VHL locus result in sporadic renal clear-cell carcinomas (RCC) as well as von Hippel-Lindau disease, which is characterized by highly vascularized tumours (reviewed in ref. 23). One target of VHL is the transcription factor hypoxia-inducible



**Figure 2 | Different strategies to target inhibitors of apoptosis (IAPs) in cancer.** IAPs are the only known endogenous inhibitors of caspases, which are essential downstream mediators of apoptotic cell-death pathways. IAPs are highly expressed in various tumours and are targeted by anticancer drugs using antisense technologies or SMAC mimetics (compounds that mimic naturally occurring IAP antagonists) to sensitize tumour cells to conventional chemotherapy. The SMAC mimetics also induce c-IAP degradation and subsequent activation of NF- $\kappa$ B. This leads to an increase in the level of tumour necrosis factor  $\alpha$  (TNF- $\alpha$ ), which kills cells through enhanced TNFR signalling. Antisense compounds against IAP members (XIAP and survivin) are being evaluated in phase I/II trials.

factor (HIF), a key mediator of oxygen homeostasis in mammalian cells. In the presence of high oxygen levels, HIF is hydroxylated at conserved prolines and recognized by VHL, which triggers ubiquitin-dependent degradation of HIF. Under hypoxic conditions, HIF escapes VHL-mediated downregulation and activates hypoxia-inducible genes, such as vascular endothelial growth factor (VEGF), erythropoietin (EPO) and glucose transporter 1 (GLUT1). These factors promote angiogenesis, increase the level of oxygen-carrying erythrocytes and boost anaerobic metabolism. Tumour cells deficient in VHL show enhanced expression of HIF target genes even under normoxic conditions and so enable high levels of vascularization and growth rates of tumours. Currently, VEGF signalling is clinically targeted in colorectal cancer, HER2-negative breast cancer and in non-squamous, non-small-cell lung cancer with the use of the monoclonal antibody bevacizumab (Avastin, Genentech/Roche), which acts as a VEGF antagonist by trapping the ligand.

### Targeting inhibitors of apoptosis

The inhibitors of apoptosis (IAPs) are a family of proteins characterized by the presence of 1–3 baculoviral IAP repeats (BIR). BIR domains confer antiapoptotic activity because they interact with caspases and inhibit their proteolytic activity<sup>24</sup>. In addition, c-IAP1 and c-IAP2 interact through their BIR domains with TRAF2 (tumour necrosis factor associated factor-2), imposing negative regulation on receptor-mediated apoptosis (Fig. 2). A large subset of the IAP family proteins also comprise a RING domain with E3 ligase activity, targeting a diverse set of substrates, including apoptotic factors and signalling molecules such as NIK, Mad1 and TRAF2. Moreover, c-IAP1, c-IAP2 and X-linked inhibitor of apoptosis (XIAP) trigger autoubiquitylation and self-degradation in response to pro-apoptotic stimuli<sup>25</sup>. Highly elevated levels of some IAPs have been reported in several types of cancer, probably due to upregulated NF- $\kappa$ B activity. Furthermore, a genetic translocation resulting in the fusion of the BIR domains of c-IAP2 to MALT1 (a ubiquitin E3 ligase) causes a type of B-cell lymphoma known as a MALT lymphoma<sup>26</sup>. BIR-1 domain-mediated self-oligomerization leads to increased MALT1 E3 ligase activity, resulting in a constitutive stimulation of NF- $\kappa$ B signalling, given that MALT1 is a mediator of T-cell-receptor-stimulated activation of NF- $\kappa$ B<sup>25,27</sup>.

Autoubiquitylation of c-IAP1 and c-IAP2 requires a mitochondrial protein called SMAC/DIABLO (second mitochondrial activator of caspases/direct IAP-binding protein with low pI), which is released from mitochondria after the induction of apoptosis (Fig. 2). This protein contains

an IBM (IAP-binding motif), which binds the BIR domains of IAPs, an event with two pro-apoptotic consequences. First, other BIR-binding proteins, such as caspases, are released from IAP-mediated inhibition; and second, the intrinsic E3 ligase activity of IAPs causes self-destruction. Different strategies have been developed to target IAPs in cancer models (Fig. 2), including downregulation of their expression by RNA interference and antisense technologies, and blocking of their functions with compounds that mimic naturally occurring IAP antagonists<sup>28</sup>. In addition, small-molecule IAP antagonists that mimic the action of SMAC have been developed. These have been shown to induce apoptosis in cancer cells by mechanisms involving NF- $\kappa$ B activation via the non-canonical pathway and the subsequent production of TNF- $\alpha$ , which then triggers cell death<sup>29,30</sup>. Preclinical data obtained from human tumour samples and animal models suggest that targeting IAP expression with antisense oligonucleotides is a promising treatment option in combination with other anticancer therapeutics<sup>31</sup>. Antisense compounds against XIAP and survivin are currently being evaluated in phase I and II trials for cancer treatment (reviewed in ref. 24).

### Deubiquitylating enzymes in cancer development

An important aspect of the ubiquitin system is its reversibility, mediated by deubiquitylating enzymes (DUBs) that recognize ubiquitylated proteins and remove their ubiquitin tags. Most human DUBs are cysteine proteases and can be divided into five subfamilies: ubiquitin-specific proteases (Usp), ubiquitin carboxy-terminal hydrolases (UCH), ovarian tumour-like proteases (OTU), JAMM/MPN metalloproteases, and Machado–Jakob-disease proteases (MJD)<sup>32</sup>. They are direct antagonists of oncogenic or tumour-suppressive E3 ligases and are increasingly being seen as potential targets in cancer therapies. Promising candidates are nuclear DUBs, including USP1, USP28 and USP44, which control processes relevant to cancer. USP1 modulates DNA-repair checkpoints by preventing the monoubiquitylation of the Fanconi anaemia complementation group D2 protein (FANCD2) and of proliferating cell nuclear antigen (PCNA)<sup>33</sup>; USP28 is overexpressed in colon and breast tumours, and by counteracting the ubiquitylation activity of SCF–Fbxw7 ligase it causes the stabilization of cyclin E1 and c-Myc<sup>34,35</sup>. USP44, however, counteracts activation of the anaphase-promoting factor APC/C by deubiquitylating Cdc20, preventing premature silencing of the spindle checkpoint<sup>36</sup>.

Other cancer-linked DUBs regulate the NF- $\kappa$ B signalling pathway. Ubiquitin ligases and DUBs implicated in the NF- $\kappa$ B pathway



are often found within multimeric complexes or even within the same polypeptide chain, which allows effective and dynamic regulation of the pathway. The most intimate relationship between DUBs and ubiquitin ligases is illustrated by the A20 protein, which harbours deubiquitylation activity in its OTU domain and an E3 ligase activity in its zinc-fingers. A20 preferentially catalyses the disassembly of Lys-63-linked ubiquitin chains from substrates such as RIP or NEMO, and promotes the formation of Lys-48-linked chains on multiple substrates, targeting them for degradation<sup>37,38</sup>.

The NF- $\kappa$ B pathway is under the control of another DUB named CYLD, a tumour-suppressor gene mutated in skin tumours (cylindromatosis). The ubiquitously expressed CYLD contains a C-terminal ubiquitin hydrolase domain that removes Lys-63-linked polyubiquitin chains from several mediators of the cytokine-induced NF- $\kappa$ B pathway (reviewed in ref. 38). Reduced or abrogated expression of CYLD has been observed in human skin cancer as well as several other tumour types, including those of the kidney, liver and uterine cervix, suggesting that CYLD has a more general role as a tumour suppressor. The analysis of CYLD-null mice led to the identification of BCL3 as a crucial *in vivo* target of CYLD for the development of cylindromatosis-related skin tumours<sup>39</sup>. BCL3 is a transcriptional co-activator that is inactive in the cytoplasm and requires polyubiquitylation to enter the nucleus. Nuclear association of BCL3 with NF- $\kappa$ B1 or NF- $\kappa$ B2 switches a transcriptional complex from being repressive into an active state and thereby promotes the expression of target genes that lead to cell proliferation.

Further understanding of the catalytic activity of DUBs, as well as their regulation and substrate specificity, will promote the development of DUB inhibitors and their use as anticancer drugs. The generation of small-molecule inhibitors targeting the ubiquitin C-terminal hydrolase (UCH-L1) has had beneficial effects in the treatment of lung tumour cells<sup>40,41</sup>. Although this initial success supports the concept of DUB inhibition as a therapeutic target, the application of DUB inhibitors in the clinic remains challenging<sup>41</sup>.

### Targeting the proteasome in tumours

Ubiquitylated proteins are targeted to the proteasome, a giant protein complex responsible for the breakdown of a large proportion of intracellular proteins<sup>42</sup>. Alterations in the proteasome are associated with several human diseases, such as cardiac dysfunction, cataract formation, neurodegenerative disorders, cachexia and rheumatoid diseases<sup>43</sup>, but defects in the proteasome have not yet been linked to cancer development. This indicates that cancer cells might exploit the proteasome for their own needs. Indeed, several antiapoptotic and proliferative signalling pathways require proteasomal activity. This applies in particular for the pro-oncogenic NF- $\kappa$ B pathway, which is activated in many types of tumour<sup>44</sup>. NF- $\kappa$ B factors are retained in an inactive state in the cytoplasm by the inhibitors of NF- $\kappa$ Bs (I $\kappa$ Bs). In order to be freed from this inhibition, I $\kappa$ Bs need to be phosphorylated, polyubiquitylated and degraded by the proteasome. The cellular effects of the boronic-acid derivative bortezomib (Velcade, Millennium Pharmaceuticals), which reversibly inhibits the active sites in the 20S proteasome, seem to favour this hypothesis. Bortezomib downregulates NF- $\kappa$ B signalling by blocking I $\kappa$ B degradation, which seems to be its prevalent mechanism of action in human tumours<sup>45</sup>. Inhibition of NF- $\kappa$ B reduces the expression of pro-inflammatory response genes and upregulates the cyclin-dependent kinase inhibitors p21<sup>Cip1</sup> and p27<sup>Kip1</sup>, resulting in increased apoptosis in tumour cells<sup>46</sup>. In addition to the effect on NF- $\kappa$ B signalling, there is evidence that, in certain cell types, bortezomib-mediated proteasome inhibition leads to endoplasmic-reticulum stress, which in turn promotes cell death<sup>47,48</sup>. Bortezomib is currently in use as an anticancer drug to treat patients with haematological tumours, such as multiple myeloma, and recently also for relapsed mantle-cell lymphoma<sup>49</sup>.

The clinical success of bortezomib as an anticancer therapy has bolstered interest in the development of a new generation of proteasome inhibitors, such as PR-171 (carfilzomib), NPI-0052 and CEP-18770 (refs 50–52). Currently available proteasome inhibitors vary in their mode of action and target distinct active sites of the 20S proteasome

(chymotrypsin-like, trypsin-like and caspase-like) for reversible inhibition, irreversible binding or covalent modification, providing a large repertoire of drugs that target the proteasome. For example, NPI-0052 irreversibly blocks the chymotrypsin- and trypsin-like sites of the proteasome, whereas bortezomib affects chymotrypsin- and caspase-like activities in a reversible manner. The anticancer efficacy of NPI-0052 is dependent on other proteasome substrates, such as FADD-caspase-8-mediated cell-death signalling<sup>53</sup>, rather than NF- $\kappa$ B signalling as in the case of bortezomib. Current preclinical and clinical studies aim to determine the efficacy of new proteasome inhibitors in haematological tumours and in a broad spectrum of solid tumours. Importantly, combinations of low doses of proteasome inhibitors, such as NPI-0052 and bortezomib, have synergistic effects on cancer cells *in vitro*<sup>54</sup>. Also, combinations of bortezomib with other antimyeloma drugs, such as melphalan, prednisone, dexamethasone and thalidomide, have been applied in phase III trials and shown promise in cases of early-diagnosed myeloma<sup>55</sup>. The outstanding challenge in the clinic is to define appropriate combinations, as well as doses, of such proteasome inhibitors, which offer therapeutic advantages over bortezomib. Moreover, development of proteasome inhibitors, with distinct substrate selectivity, improved bio-availability and lower toxicity, may open the door to widespread usage in solid tumours. An example of the next generation of proteasome inhibitor is argyirin A, which was identified in screens for compounds that stabilize the cell-cycle inhibitor p27<sup>Kip1</sup>. The antitumour activities of argyirin A were shown to depend on the expression of p27<sup>Kip1</sup>, and resistance to argyirin A was directly linked to the loss of p27<sup>Kip1</sup> (ref. 56).

### Non-degradative ubiquitylation and cancer pathogenesis

Ubiquitylation is also involved in non-proteolytic functions in tumour cells, including the assembly of pro-oncogenic signalling complexes (NF- $\kappa$ B), shuttling of proto-oncogenes or tumour suppressors (such as p53 and PTEN) between the cytosol and the nucleus, and recruitment of tumour-promoting complexes to subcellular compartments such as nuclear foci (reviewed in refs 2, 4, 44). These functions are achieved by the formation of atypical ubiquitin chains, linked through lysines at positions other than 48 or 63, such as Lys 6, Lys 11, Lys 27, Lys 29 or Lys 33, or by a single ubiquitin molecule attached to distinct substrates<sup>7</sup>. Such dynamic ubiquitin networks are well documented for cellular responses to DNA damage. To avoid deleterious mutations, the cell needs to rapidly detect DNA damage, identify the nature of the lesion and recruit the appropriate repair machinery. The RING-type E3 ligases RNF8 and BRCA1 (from breast cancer susceptibility gene 1) have been recognized as key players in the repair of DSBs caused by ionizing radiation (reviewed in ref. 5; Fig. 3). In cooperation with the E2 enzyme UBC13, RNF8 assembles Lys-63-linked chains on histones H2A and H2AX. The non-degradative ubiquitin tags provide a crucial signal for the assembly of the late effector complex, which contains BRCA1 along with RAP80 and other proteins. RAP80 contains two ubiquitin-interacting motifs (UIMs), which bind preferentially to Lys-63-linked ubiquitin chains and are required for the recruitment of the BRCA1 complex and the formation of nuclear foci, essential for the repair of the damaged DNA (reviewed in ref. 2). Mutations in the *BRCA1* locus that lead to the expression of a truncated or E3-inactive BRCA1 protein are strongly linked to the development of ovarian and breast cancer<sup>57</sup>, but alterations in the genes encoding RAP80 or RNF8 have not yet been associated with cancer.

Another way to modulate oncogenic signalling pathways is to alter the subcellular localization of their key components. This often involves monoubiquitylation as a trigger for the relocation of a protein. In the case of the tumour-suppressor protein p53, the molecular requirements for ubiquitylation have been particularly well studied (reviewed in ref. 58). The attachment of monoubiquitin to the C terminus of p53 seems to induce a conformational change that exposes a nuclear export signal (NES)<sup>59,60</sup>. In the cytosol, p53 is believed to promote apoptotic events directly, independent of transcription<sup>61</sup>. The functional relevance of ubiquitin chains and monoubiquitin on p53 may differ dramatically: whereas polyubiquitylation destroys the

gatekeeper, monoubiquitylation causes cell death by relocating the gatekeeper to the cytosol.

The tumour suppressor PTEN (phosphatase and tensin homologue) is also subject to monoubiquitin- and polyubiquitin-dependent regulation<sup>62,63</sup>. PTEN is a lipid phosphatase that negatively regulates the phosphatidylinositol 3-kinase (PI3K)/AKT signalling cascade and is one of the genes most frequently inactivated in human cancers (reviewed in ref. 64). The physiological relevance of the monoubiquitylation-dependent nuclear translocation of PTEN is illustrated by mutations of the ubiquitin-acceptor lysines, Lys 289 and Lys 13, which are affected in Cowden syndrome (a cancer-susceptibility syndrome) and sporadic brain tumours, respectively. In both cases, mutated PTEN retains its phosphatase activity as well as membrane localization, but it is excluded

from the nucleus. Moreover, the failure of PTEN nuclear localization has been associated with advanced stages of tumorigenesis, and is often followed by a total loss of PTEN expression, a feature strongly associated with late stages of disease<sup>65</sup>. The HECT-type E3 ligase Nedd4-1 has been reported to mediate both monoubiquitylation and polyubiquitin-dependent proteasomal removal of PTEN *in vitro*<sup>63</sup>. However, the regulation of PTEN ubiquitylation is more complex *in vivo*, as the deletion of Nedd4-1 is dispensable for the regulation of PTEN stability and subcellular localization in mice<sup>66</sup>. It is therefore likely that multiple ubiquitin ligases act in a cell- and compartment-specific manner to control ubiquitin-dependent regulation of PTEN *in vivo*. Herpesvirus-associated ubiquitin-specific protease (HAUSP) was recently found to deubiquitylate PTEN, leading to its exclusion from the nucleus<sup>67</sup> (Fig. 4a). This effect was counteracted by PML, which inhibits HAUSP activity through DAXX (a transcriptional repressor). PML is the major structural component of PML nuclear bodies and in acute promyelocytic leukaemia (APL) is fused to the retinoic acid receptor RAR- $\alpha$ . This fusion event not only disrupts PML function, but also seems to cause a cytosolic accumulation of PTEN, which might be explained by aberrant activation of HAUSP (Fig. 4a). Patients with APL are commonly treated with either arsenic trioxide or all-*trans* retinoic acid, which induce degradation of PML-RAR- $\alpha$ . Both substances were also found to increase monoubiquitylation and nuclear localization of PTEN.

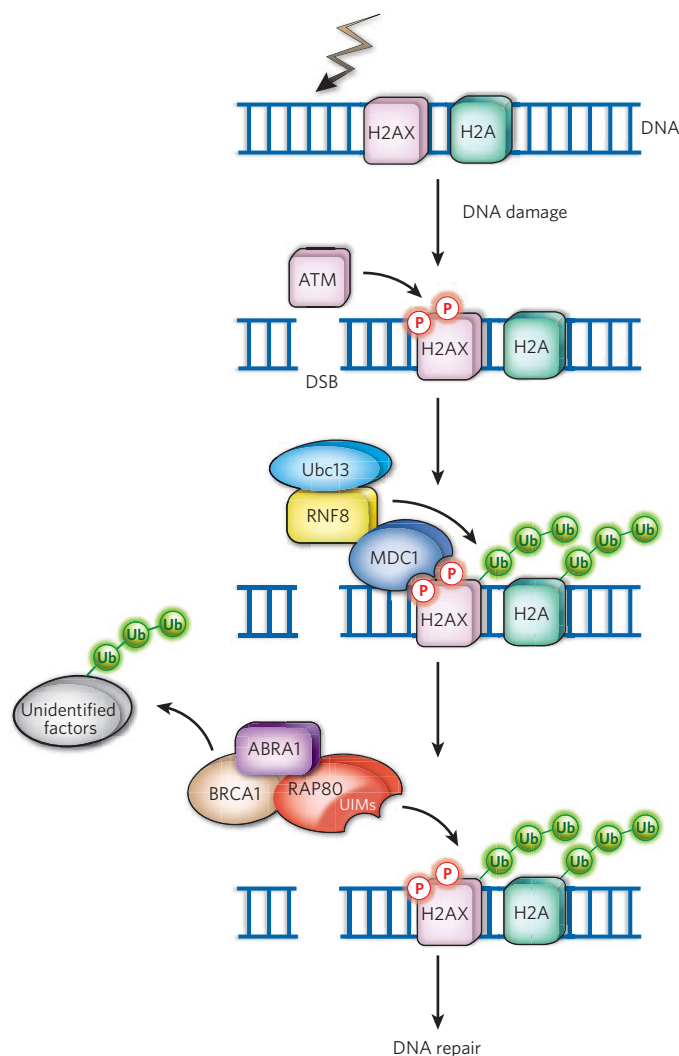
### Interfering with recognition of ubiquitin

Ubiquitylation generates a broad spectrum of distinct signals with specific functional consequences for the modified protein. Ubiquitin-binding domains (UBDs) are small modular domains with the capacity to bind and distinguish between the different types of ubiquitin modifications (monoubiquitin and polyubiquitin chains of different linkage types and length)<sup>7,68</sup>. More than 20 families of UBD have been identified. Their specific roles can be exemplified by regulation of the double-strand DNA repair pathways (the UIMs in RAP80), translesion synthesis (TLS) pathways (the UBM and UBZ domains in TLS polymerases)<sup>69</sup> and the activation of protein kinases in the NF- $\kappa$ B pathway (UBDs in NEMO and TAB, reviewed in ref. 70).

One of the most challenging therapeutic approaches will be to interfere with ubiquitin-UBD interactions in cells (Fig. 1c). This is a difficult task, as the hydrophobic binding surface between ubiquitin and UBDs is typically flat and the affinity of interaction is low. However, in many cases the low affinity is compensated by the presence of multiple interaction sites within the protein complex. Small-molecule inhibitors called ubistatins are the first compounds to be identified that interfere with ubiquitin-UBD interactions. Ubistatins act by binding to the hydrophobic interface of Lys-48-linked polyubiquitin<sup>71</sup>, preventing the recognition of ubiquitylated substrates by UBDs present in the proteasome receptors RPN10 and RPN13, as well as the shuttling receptor RAD23 that targets the same surface<sup>72</sup> (Fig. 1c). Although the clinical use of ubistatins is ruled out by their inability to penetrate cell membranes, they show that ubiquitin-UBD interactions can be efficiently disrupted by small compounds. As well as the need to generate bioactive substances, another challenging step will be to generate molecules that target the UBD of a given effector protein. This would allow the specific manipulation of distinct cellular processes, rather than a general interference with the recognition of ubiquitylated proteins.

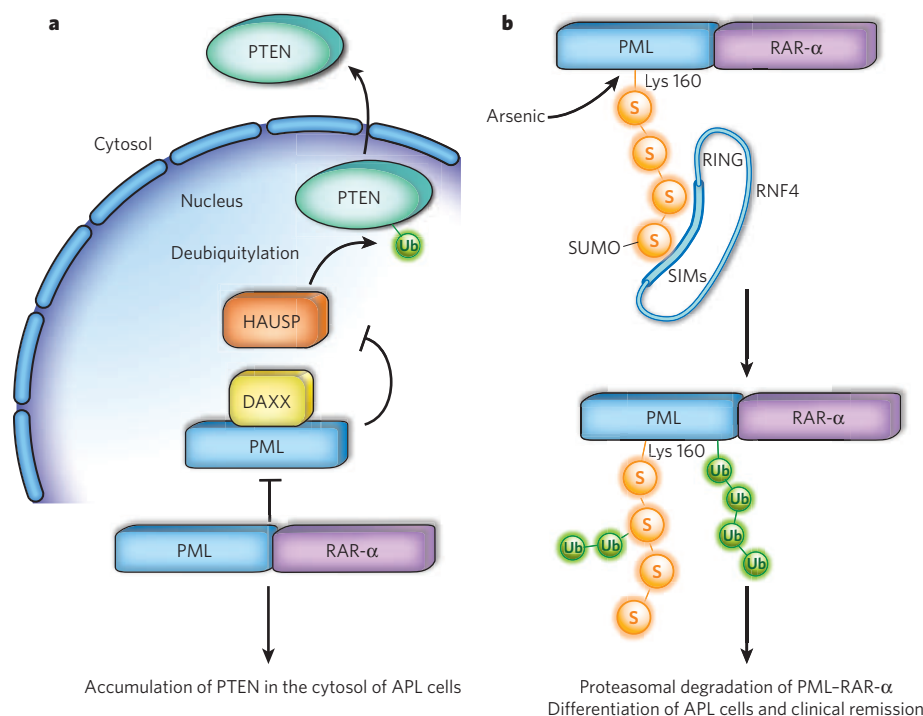
### SUMO and cancer development

Ubiquitin is the first of a growing family of ubiquitin-like (UBL) proteins that is defined by a common structure (the ubiquitin globular  $\beta$ -grasp fold) rather than by sequence homology. In addition, all UBLs have a highly conserved C-terminal glycine (essential for conjugation) and a similar conjugation system comprising E1, E2 and E3 enzymes. SUMO (small ubiquitin-like modifier) is the best-characterized of a dozen UBL modifiers identified in mammals. It seems to be mainly involved in the regulation of nuclear processes, such as transcriptional repression, DNA replication, DNA damage repair and the formation of PML nuclear bodies, as well as in the nucleo-cytoplasmic shuttling



**Figure 3 | Non-degradative ubiquitin modifications have a role in DNA repair.** Cellular insults such as radiation can damage DNA by causing double-strand breaks (DSBs). The checkpoint kinase ATM is recruited to the site of damage and induces phosphorylation of histone 2AX (H2AX). This in turn attracts MDC1 and the E3 ligase RNF8, which cooperates with the E2-conjugating enzyme UBC13 to assemble Lys-63-linked ubiquitin chains on the histone proteins. This type of ubiquitylation does not lead to degradation of the modified proteins but is instead required for the assembly of a protein complex that deals with the DNA damage. A crucial component of this complex is RAP80, which contains two ubiquitin-interacting motifs (UIMs) that specifically recognize the Lys-63-linked chains on histone H2A. Along with other proteins, RAP80 mediates the recruitment of the E3 ligase BRCA1 to the site of DNA damage. Mutations in the *BRCA1* gene are associated with ovarian and breast cancer, although the identities of its ubiquitylation substrates remain elusive.





**Figure 4 | Targeting the ubiquitin system in the nucleus.** **a**, The nuclear localization of the tumour suppressor PTEN is dependent on its monoubiquitylation. Deubiquitylation by HAUSP causes PTEN to accumulate in the cytosol. HAUSP activity is inhibited by PML protein. In patients with acute promyelocytic leukaemia (APL), PML function is disrupted by the oncogenic PML-RAR- $\alpha$  fusion protein, blocking the inhibition of HAUSP and causing the removal of PTEN from the nucleus. **b**, RNF4 mediates the arsenic-induced degradation of PML-RAR- $\alpha$  (as well as the native PML protein). Arsenic stimulates the attachment of chains of small ubiquitin-like modifiers (SUMOs) on Lys 160, which recruit RNF4 via its SUMO-interacting motif (SIM) domains. RNF4 then mediates polyubiquitylation and the subsequent proteasome-dependent degradation of PML-RAR- $\alpha$ . Treatment of patients with APL using arsenic trioxide therefore results in the differentiation of leukaemic cells and remission of the disease.

of proteins. Differing from the ubiquitin pathway, SUMOylation uses only a single E2 enzyme, UBC9, which interacts directly with substrates that are often modified on a lysine located within a specific consensus sequence (reviewed in ref. 73). UBC9 is overexpressed in several human tumours, including lung adenocarcinoma, ovarian carcinoma and melanoma. It occupies a key position within the SUMO system and is an attractive target for therapeutic intervention. Several approaches to target UBC9 functions are in experimental phases, including inhibition of the active site, interference with binding to the E1 enzyme, and blocking the binding to target proteins.

Like the ubiquitin system, the physiological outcome of SUMOylation is commonly mediated by proteins containing SUMO-interacting motifs (SIMs). One example of such proteins is the ubiquitin E3 ligase RNF4 (also called SNURF), which contains four SIMs and links the SUMO and ubiquitin systems. RNF4 has an important role in APL, as it recognizes SUMOylated PML as well as the APL-associated oncogenic fusion protein PML-RAR- $\alpha$ , and triggers their polyubiquitin-dependent proteasomal degradation<sup>74,75</sup> (Fig. 4b). PML is the essential structural component of PML nuclear bodies, which are dynamic proteinaceous structures implicated in various cellular functions such as apoptosis, inhibition of proliferation, maintenance of genomic stability, and antiviral responses. Intriguingly, PML nuclear bodies are lost in APL as well as in several solid tumours, suggesting a tumour-protective function.

Degradation of PML-RAR- $\alpha$  (as well as the native PML protein) can be stimulated by arsenic trioxide, which enhances the attachment of SUMO-1 and SUMO-2 chains to PML-RAR- $\alpha$  via its four SIM domains. RNF4 specifically binds polySUMO-2 chains and thereby mediates the recruitment of RNF4 into PML-nuclear bodies upon arsenic treatment (Fig. 4b). RNF4 ubiquitylates the PML moiety, as well as the SUMO-2 chain of the fusion protein<sup>74,75</sup>. Subsequent proteasomal degradation of PML-RAR- $\alpha$  results in differentiation of leukaemic cells and clinical remissions. Arsenic trioxide, the infamous compound in Madame Bovary's poison, is currently used in clinics as an efficient therapeutic agent for the treatment of patients with APL. More generally, the mechanism of action of arsenic trioxide points to the possibility of designing drugs that target proteins to the ubiquitin system.

### Future challenges

A large body of evidence, coming mostly from studies in cultured cells, has provided molecular explanations of how alterations in the

ubiquitin system can contribute to the aetiology of different types of tumour. In many cases these data have already been validated in clinical trials, allowing the development of strategies to target the ubiquitin system. Yet the results obtained in cultured-cell studies do not always reflect the findings later gained in *in vivo* models<sup>62,63,66</sup>; it will be necessary to analyse patient tumour samples and established mouse models. Complicating the situation further, ubiquitin signalling pathways often act alongside other oncogenic cascades during cell transformation and metastasis. To understand the dynamics and complexity of such events *in vivo*, it will be necessary to initiate comprehensive proteomic and genomic studies with primary tumour materials from a large number of patients.

Interfering with ubiquitin pathways, for example using the proteasome inhibitor bortezomib, is a valid strategy for the development of anticancer drugs. Several approaches are being developed to further interfere with aberrant components of the ubiquitin system in tumours, including inhibition of ubiquitin conjugases and ligases (Fig. 1a, b), blocking substrate recognition, preventing binding of UBDs to ubiquitin and designing new proteasome inhibitors (Fig. 1c). In some cases the outcome has been disappointing, notably efforts to develop inhibitors of specific ubiquitin ligases, but attempts to generate new proteasome inhibitors have been more fruitful. The use of antisense compounds and strategies to block protein-protein interactions both hold promise in current clinical studies, as exemplified by antagonists of IAPs (Fig. 2; reviewed in ref. 24). UBL networks will also be explored therapeutically as they are involved in processes that may underlie tumour resistance, such as autophagy. If clinical progress is made, this will expand and diversify the range of anticancer therapies needed for the more efficient treatment of cancer patients in the future.

1. Hershko, A. & Ciechanover, A. The ubiquitin system. *Annu. Rev. Biochem.* **67**, 425-479 (1998).
2. Bennett, E. J. & Harper, J. W. DNA damage: ubiquitin marks the spot. *Nature Struct. Mol. Biol.* **15**, 20-22 (2008).
3. Katzmann, D. J., Odorizzi, G. & Emr, S. D. Receptor downregulation and multivesicular-body sorting. *Nature Rev. Mol. Cell Biol.* **3**, 893-905 (2002).
4. Haglund, K. & Dikic, I. Ubiquitylation and cell signaling. *EMBO J.* **24**, 3353-3359 (2005).
5. Mukhopadhyay, D. & Riezman, H. Proteasome-independent functions of ubiquitin in endocytosis and signaling. *Science* **315**, 201-205 (2007).
6. Pickart, C. M. Mechanisms underlying ubiquitination. *Annu. Rev. Biochem.* **70**, 503-533 (2001).
7. Ikeda, F. & Dikic, I. Atypical ubiquitin chains: new molecular signals. *EMBO Rep.* **9**, 536-542 (2008).

8. Hoeller, D., Hecker, C. M. & Dikic, I. Ubiquitin and ubiquitin-like proteins in cancer pathogenesis. *Nature Rev. Cancer* **6**, 776–788 (2006).
9. Bernassola, F., Karin, M., Ciechanover, A. & Melino, G. The HECT family of E3 ubiquitin ligases: multiple players in cancer development. *Cancer Cell* **14**, 10–21 (2008).
10. Nakayama, K. I. & Nakayama, K. Ubiquitin ligases: cell-cycle control and cancer. *Nature Rev. Cancer* **6**, 369–381 (2006).
11. Bond, G. L., Hu, W. & Levine, A. J. MDM2 is a central node in the p53 pathway: 12 years and counting. *Curr. Cancer Drug Targets* **5**, 3–8 (2005).
12. Vousden, K. H. & Prives, C. p53 and prognosis: new insights and further complexity. *Cell* **120**, 7–10 (2005).
13. Cardozo, T. & Pagano, M. Wrenches in the works: drug discovery targeting the SCF ubiquitin ligase and APC/C complexes. *BMC Biochem.* **8** (suppl. 1), S9 (2007).
14. Vassilev, L. T. MDM2 inhibitors for cancer therapy. *Trends Mol. Med.* **13**, 23–31 (2007).
15. Guedat, P. & Colland, F. Patented small molecule inhibitors in the ubiquitin proteasome system. *BMC Biochem.* **8** (suppl. 1), S14 (2007).
16. Hjerpe, R. & Rodriguez, M. S. Alternative UPS drug targets upstream the 26S proteasome. *Int. J. Biochem. Cell Biol.* **40**, 1126–1140 (2008).
17. Yang, Y. *et al.* Small molecule inhibitors of HDM2 ubiquitin ligase activity stabilize and activate p53 in cells. *Cancer Cell* **7**, 547–559 (2005).  
**This is the first report about the development of inhibitors that target the active site of a RING-type E3. The identified compound, HLI98, shows specificity for MDM2 and allows the activation of p53-dependent transcription in cells.**
18. VanderBorght, A. *et al.* Effect of an *hdm-2* antagonist peptide inhibitor on cell cycle progression in p53-deficient H1299 human lung carcinoma cells. *Oncogene* **25**, 6672–6677 (2006).
19. Brooks, C. L. & Gu, W. p53 ubiquitination: Mdm2 and beyond. *Mol. Cell* **21**, 307–315 (2006).
20. Issaeva, N. *et al.* Small molecule RITA binds to p53, blocks p53–HDM-2 interaction and activates p53 function in tumors. *Nature Med.* **10**, 1321–1328 (2004).  
**The authors identify RITA as a potent inhibitor of p53 degradation. RITA induced p53-dependent apoptosis in various tumour cells, and showed significant antitumour effects in vivo.**
21. Frescas, D. & Pagano, M. Deregulated proteolysis by the F-box proteins SKP2 and  $\beta$ -TrCP: tipping the scales of cancer. *Nature Rev. Cancer* **8**, 438–449 (2008).
22. Hanahan, D. & Folkman, J. Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis. *Cell* **86**, 353–364 (1996).
23. Kaelin, W. G. Jr. The von Hippel–Lindau tumor suppressor protein and clear cell renal carcinoma. *Clin. Cancer Res.* **13**, 6805–6845 (2007).
24. Srinivasula, S. M. & Ashwell, J. D. IAPs: what's in a name? *Mol. Cell* **30**, 123–135 (2008).
25. Vaux, D. L. & Silke, J. IAPs, RINGs and ubiquitylation. *Nature Rev. Mol. Cell Biol.* **6**, 287–297 (2005).
26. Dierlamm, J. *et al.* The apoptosis inhibitor gene *API2* and a novel 18q gene, *MLT*, are recurrently rearranged in the t(11;18)(q21;q21) associated with mucosa-associated lymphoid tissue lymphomas. *Blood* **93**, 3601–3609 (1999).
27. Noels, H. *et al.* A novel TRAF6 binding site in MALT1 defines distinct mechanisms of NF- $\kappa$ B activation by API2–MALT1 fusions. *J. Biol. Chem.* **282**, 10180–10189 (2007).
28. Wright, C. W. & Duckett, C. S. Reawakening the cellular death program in neoplasia through the therapeutic blockade of IAP function. *J. Clin. Invest.* **115**, 2673–2678 (2005).
29. Varfolomeev, E. *et al.* c-IAP1 and c-IAP2 are critical mediators of TNF $\alpha$ -induced NF- $\kappa$ B activation. *J. Biol. Chem.* **283**, 24295–24299 (2008).
30. Vince, J. E. *et al.* IAP antagonists target cIAP1 to induce TNF $\alpha$ -dependent apoptosis. *Cell* **131**, 682–693 (2007).
31. Cummings, J. *et al.* Method validation and preliminary qualification of pharmacodynamic biomarkers employed to evaluate the clinical efficacy of an antisense compound (AEG35156) targeted to the X-linked inhibitor of apoptosis protein XIAP. *Br. J. Cancer* **95**, 42–48 (2006).
32. Nijman, S. M. *et al.* A genomic and functional inventory of deubiquitinating enzymes. *Cell* **123**, 773–786 (2005).
33. Nijman, S. M. *et al.* The deubiquitinating enzyme USP1 regulates the Fanconi anemia pathway. *Mol. Cell* **17**, 331–339 (2005).
34. Popov, N., Herold, S., Llamazares, M., Schulein, C. & Eilers, M. Fbw7 and Usp28 regulate Myc protein stability in response to DNA damage. *Cell Cycle* **6**, 2327–2331 (2007).
35. Popov, N. *et al.* The ubiquitin-specific protease USP28 is required for MYC stability. *Nature Cell Biol.* **9**, 765–774 (2007).
36. Stegmeier, F. *et al.* Anaphase initiation is regulated by antagonistic ubiquitination and deubiquitination activities. *Nature* **446**, 876–881 (2007).
37. Wertz, I. E. *et al.* De-ubiquitination and ubiquitin ligase domains of A20 downregulate NF- $\kappa$ B signalling. *Nature* **430**, 694–699 (2004).  
**This paper describes the molecular mechanism of ubiquitin-chain editing by A20, a single molecule containing E3 ligase and deubiquitylating activities.**
38. Sun, S. C. Deubiquitylation and regulation of the immune response. *Nature Rev. Immunol.* **8**, 501–511 (2008).
39. Massoumi, R., Chmielarska, K., Hennecke, K., Pfeifer, A. & Fassler, R. Cylid inhibits tumor cell proliferation by blocking Bcl-3-dependent NF- $\kappa$ B signaling. *Cell* **125**, 665–677 (2006).
40. Liu, Y. *et al.* Discovery of inhibitors that elucidate the role of UCH-L1 activity in the H1299 lung cancer cell line. *Chem. Biol.* **10**, 837–846 (2003).
41. Love, K. R., Catic, A., Schlieker, C. & Ploegh, H. L. Mechanisms, biology and inhibitors of deubiquitinating enzymes. *Nature Chem. Biol.* **3**, 697–705 (2007).
42. Rubin, D. M. & Finley, D. The proteasome: a protein-degrading organelle? *Curr. Biol.* **5**, 854–858 (1995).
43. Dahlmann, B. Role of proteasomes in disease. *BMC Biochem.* **8** (suppl. 1), S3 (2007).
44. Karin, M. Nuclear factor- $\kappa$ B in cancer development and progression. *Nature* **441**, 431–436 (2006).
45. Adams, J. The development of proteasome inhibitors as anticancer drugs. *Cancer Cell* **5**, 417–421 (2004).
46. Ling, Y. H. *et al.* Mechanisms of proteasome inhibitor PS-341-induced G<sub>2</sub>–M-phase arrest and apoptosis in human non-small cell lung cancer cell lines. *Clin. Cancer Res.* **9**, 1145–1154 (2003).
47. Meister, S. *et al.* Extensive immunoglobulin production sensitizes myeloma cells for proteasome inhibition. *Cancer Res.* **67**, 1783–1792 (2007).
48. Gu, H., Chen, X., Gao, G. & Dong, H. Caspase-2 functions upstream of mitochondria in endoplasmic reticulum stress-induced apoptosis by bortezomib in human myeloma cells. *Mol. Cancer Ther.* **7**, 2298–2307 (2008).
49. Tobinai, K. Proteasome inhibitor, bortezomib, for myeloma and lymphoma. *Int. J. Clin. Oncol.* **12**, 318–326 (2007).
50. Dorsey, B. D. *et al.* Discovery of a potent, selective, and orally active proteasome inhibitor for the treatment of cancer. *J. Med. Chem.* **51**, 1068–1072 (2008).
51. Piva, R. *et al.* CEP-18770: a novel, orally active proteasome inhibitor with a tumor-selective pharmacologic profile competitive with bortezomib. *Blood* **111**, 2765–2775 (2008).
52. Nalepa, G., Rolfe, M. & Harper, J. W. Drug discovery in the ubiquitin–proteasome system. *Nature Rev. Drug Discov.* **5**, 596–613 (2006).
53. Miller, C. P. *et al.* NPI-0052, a novel proteasome inhibitor, induces caspase-8 and ROS-dependent apoptosis alone and in combination with HDAC inhibitors in leukemia cells. *Blood* **110**, 267–277 (2007).
54. Chauhan, D. *et al.* Combination of proteasome inhibitors bortezomib and NPI-0052 trigger *in vivo* synergistic cytotoxicity in multiple myeloma. *Blood* **111**, 1654–1664 (2008).
55. Richardson, P. G. *et al.* Bortezomib in the front-line treatment of multiple myeloma. *Expert Rev. Anticancer Ther.* **8**, 1053–1072 (2008).
56. Nicleleit, I. *et al.* Argyrin A reveals a critical role for the tumor suppressor protein p27<sup>Wip1</sup> in mediating antitumor activities in response to proteasome inhibition. *Cancer Cell* **14**, 23–35 (2008).
57. Hashizume, R. *et al.* The RING heterodimer BRCA1–BARD1 is a ubiquitin ligase inactivated by a breast cancer-derived mutation. *J. Biol. Chem.* **276**, 14537–14540 (2001).
58. Brooks, C. L., Li, M. & Gu, W. Monoubiquitination: the signal for p53 nuclear export? *Cell Cycle* **3**, 436–438 (2004).
59. Nie, L., Sasaki, M. & Maki, C. G. Regulation of p53 nuclear export through sequential changes in conformation and ubiquitination. *J. Biol. Chem.* **282**, 14616–14625 (2007).
60. Carter, S., Bischof, O., Dejean, A. & Vousden, K. H. C-terminal modifications regulate MDM2 dissociation and nuclear export of p53. *Nature Cell Biol.* **9**, 428–435 (2007).
61. Moll, U. M., Wolff, S., Speidel, D. & Deppert, W. Transcription-independent pro-apoptotic functions of p53. *Curr. Opin. Cell Biol.* **17**, 631–636 (2005).
62. Trotman, L. C. *et al.* Ubiquitination regulates PTEN nuclear import and tumor suppression. *Cell* **128**, 141–156 (2007).
63. Wang, X. *et al.* NEDD4-1 is a proto-oncogenic ubiquitin ligase for PTEN. *Cell* **128**, 129–139 (2007).
64. Cully, M., You, H., Levine, A. J. & Mak, T. W. Beyond PTEN mutations: the PI3K pathway as an integrator of multiple inputs during tumorigenesis. *Nature Rev. Cancer* **6**, 184–192 (2006).
65. Gimm, O. *et al.* Differential nuclear and cytoplasmic expression of PTEN in normal thyroid tissue, and benign and malignant epithelial thyroid tumors. *Am. J. Pathol.* **156**, 1693–1700 (2000).
66. Fouladkou, F. *et al.* The ubiquitin ligase Nedd4-1 is dispensable for the regulation of PTEN stability and localization. *Proc. Natl Acad. Sci. USA* **105**, 8585–8590 (2008).
67. Song, M. S. *et al.* The deubiquitylation and localization of PTEN are regulated by a HAUSP–PML network. *Nature* **455**, 813–817 (2008).
68. Hurley, J. H., Lee, S. & Prag, G. Ubiquitin-binding domains. *Biochem. J.* **399**, 361–372 (2006).
69. Bienko, M. *et al.* Ubiquitin-binding domains in Y-family polymerases regulate translesion synthesis. *Science* **310**, 1821–1824 (2005).
70. Chen, Z. J. Ubiquitin signalling in the NF- $\kappa$ B pathway. *Nature Cell Biol.* **7**, 758–765 (2005).
71. Verma, R. *et al.* Ubistatins inhibit proteasome-dependent degradation by binding the ubiquitin chain. *Science* **306**, 117–120 (2004).  
**Ubistatins are the first example of small molecules interfering with the recognition of specific ubiquitin chains by ubiquitin receptors.**
72. Husnjak, K. *et al.* Proteasome subunit Rpn13 is a novel ubiquitin receptor. *Nature* **453**, 481–488 (2008).  
**This paper shows that Rpn13 is a second proteasomal ubiquitin receptor that links the recognition of ubiquitylated proteins with the disassembly of ubiquitin chains by Uch37.**
73. Tatham, M. H. *et al.* Unique binding interactions among Ubc9, SUMO and RanBP2 reveal a mechanism for SUMO paralogue selection. *Nature Struct. Mol. Biol.* **12**, 67–74 (2005).
74. Tatham, M. H. *et al.* RNF4 is a poly-SUMO-specific E3 ubiquitin ligase required for arsenic-induced PML degradation. *Nature Cell Biol.* **10**, 538–546 (2008).
75. Lallemand-Breitenbach, V. *et al.* Arsenic degrades PML or PML–RAR $\alpha$  through a SUMO-triggered RNF4/ubiquitin-mediated pathway. *Nature Cell Biol.* **10**, 547–555 (2008).  
**References 74 and 75 describe the SUMO-dependent ubiquitylation of a protein and explains how arsenic-induced SUMOylation leads to degradation of the PML–RAR- $\alpha$  oncogene.**

**Acknowledgements** We thank A. Ciechanover, R. Deshaies, M. Pagano, K. Rajalingam, D. Vucic and members of the Dikic laboratory for critical reading of the manuscript. We apologize to those investigators whose contributions are not described here because of space limitations. D.H. is supported by a European Molecular Biology Organization (EMBO) long-term fellowship. I.D. acknowledges support from the German Research Foundation (DFG) and the Cluster of Excellence 'Macromolecular Complexes' (Goethe University Frankfurt).

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Correspondence should be addressed to I.D. ([dikic@biochem2.uni-frankfurt.de](mailto:dikic@biochem2.uni-frankfurt.de)).



# The ESCRT machinery in endosomal sorting of ubiquitylated membrane proteins

Camilla Raiborg<sup>1,2</sup> & Harald Stenmark<sup>1,2</sup>

**Selective trafficking of membrane proteins to lysosomes for destruction is required for proper cell signalling and metabolism. Ubiquitylation aids this process by specifying which proteins should be transported to the lysosome lumen by the multivesicular endosome pathway. The endosomal sorting complex required for transport (ESCRT) machinery sorts cargo labelled with ubiquitin into invaginations of endosome membranes. Then, through a highly conserved mechanism also used in cytokinesis and viral budding, it mediates the breaking off of the cargo-containing intraluminal vesicles from the perimeter membrane. The involvement of the ESCRT machinery in suppressing diseases such as cancer, neurodegeneration and infections underscores its importance to the cell.**

The first known function of ubiquitin was to mark cytosolic proteins for proteasomal degradation, and it is now evident that this small protein also serves to label membrane proteins for lysosomal destruction. Ubiquitin-mediated degradation of membrane proteins is crucial for quality control in the cell and for the attenuation of receptor-mediated signalling pathways. Misfolded proteins in the plasma membrane, as well as activated growth-factor, hormone and cytokine receptors, are brought inside the cell by endocytosis and delivered to lysosomes by multivesicular endosomes (MVEs) (Fig. 1), often referred to as 'multivesicular bodies' (MVBs)<sup>1</sup>. Proteasomal degradation is signalled by polyubiquitin conjugation through lysine residues at position 48 (Lys-48-linked), whereas multiple monoubiquitylation and Lys-63-linked polyubiquitylation mediate protein degradation in the lysosome by functioning as sorting signals in the endosome membrane. Ubiquitylated proteins arriving at this destination are sorted into intraluminal vesicles (ILVs) and are not recycled back to the plasma membrane, transported retrogradely to the secretory pathway, or retained in the limiting endosome membrane (Fig. 1).

The molecular basis of ubiquitin-dependent endosomal sorting is now beginning to emerge. In particular, the isolation of vacuolar protein sorting (*vps*) mutants in the yeast *Saccharomyces cerevisiae* has identified a conserved mechanism for ubiquitin-dependent sorting of membrane proteins from the limiting membranes into the ILVs of endosomes<sup>2</sup>. When the resulting MVEs fuse with lysosomes (or, in yeast, the vacuole), the ILV membrane is degraded by lipases and its transmembrane content processed by lysosomal proteases<sup>1,2</sup>. The endosomal sorting of such cargoes is initiated with their ubiquitylation by specific E3 ubiquitin ligases (reviewed in ref. 3), which often begins at the plasma membrane and continues on endosomal membranes. The ubiquitylation of some membrane proteins may actually serve as internalization signal<sup>4–6</sup>, although there are many other ways of triggering the endocytosis of specific proteins.

In the endosome membrane, the ubiquitylated cargo is captured by the endosomal sorting complex for transport (ESCRT) machinery<sup>2</sup>. This conserved machinery performs three distinct but connected functions: first, it recognizes ubiquitylated cargoes and prevents their recycling and retrograde trafficking; second, it deforms the endosomal membrane,

allowing cargo to be sorted into endosomal invaginations; third, it catalyses the final abscission (breaking off) of the endosomal invaginations, forming ILVs that contain the sorted cargo. How these three tasks are accomplished is currently being revealed. The emerging picture is that discrete ubiquitin-binding subcomplexes contribute to cargo sorting and the nucleation of a multimeric subcomplex that mediates membrane deformation and vesicle abscission. Here we review how ESCRTs mediate the sorting of ubiquitylated membrane proteins for lysosomal destruction, and the physiological importance of this pathway.

## Composition and recruitment of ESCRTs

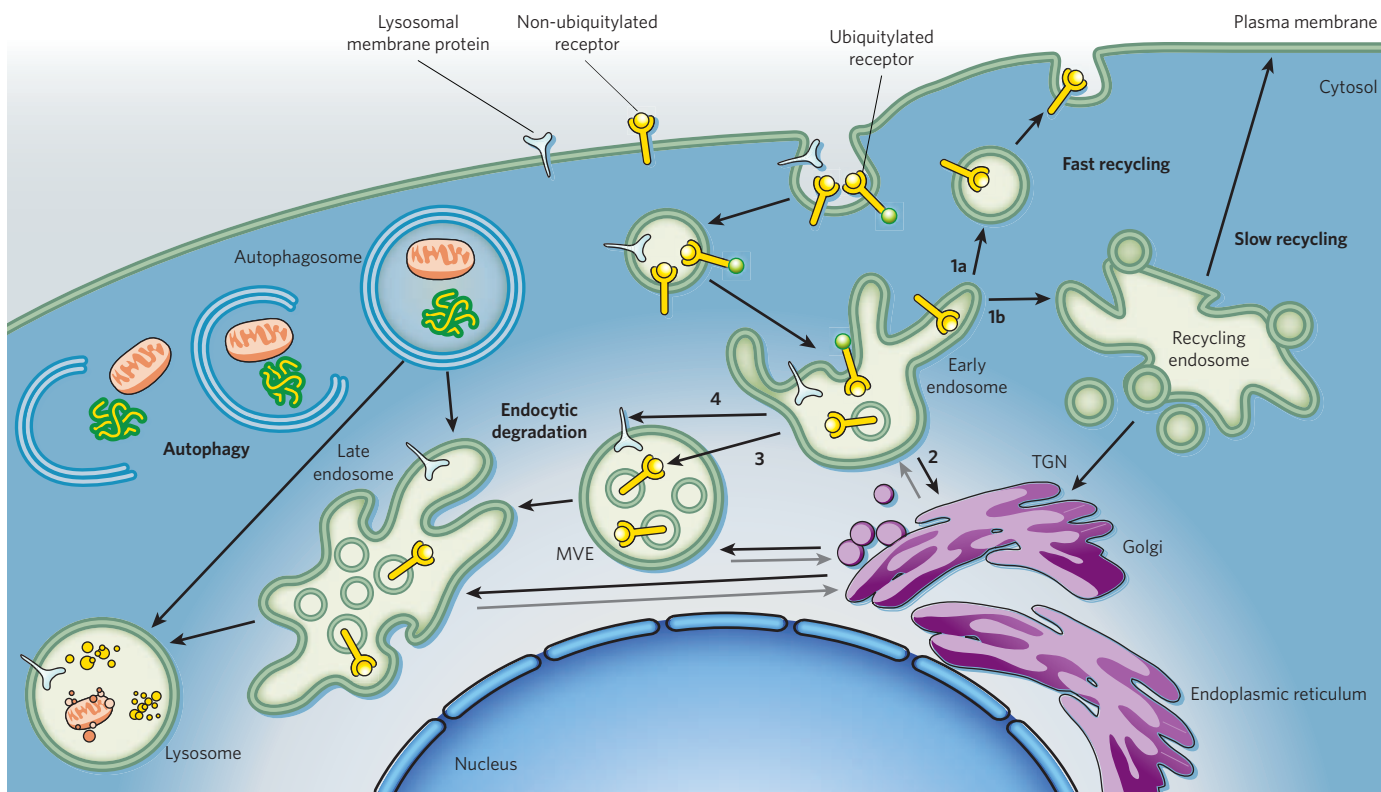
The ESCRT machinery consists of four complexes, ESCRT-0, -I, -II and -III, plus several accessory components (Fig. 2 and Table 1), whose order of recruitment and function has been unravelled by genetic and biochemical approaches<sup>7–12</sup>. Structural information is available for all ESCRTs, and the detailed structural biology of the ESCRTs has recently been described in excellent reviews<sup>13,14</sup>.

### ESCRT-0

This complex consists of the subunits Hrs and STAM (Vps27 and Hse1 in yeast). It was not originally classified as a member of the ESCRTs, but biochemical, cell-biological and genetic evidence suggests that it functions to recruit ESCRT-I (refs 10–12), so it seems logical to categorize it as an ESCRT<sup>14</sup>. The structure of the ESCRT-0 core complex consists of two intertwined GAT domains, each consisting of two helices from one subunit and one from the other<sup>15</sup>. An antiparallel coiled-coil connects the two GAT domains. As well as the occurrence of cryptic GAT domains, there are several similarities between the two subunits of ESCRT-0. They both contain an amino-terminal VHS (Vps27, Hrs and STAM) domain of unknown function, as well as ubiquitin- and clathrin-binding domains<sup>16–18</sup>. What is special to the Hrs subunit is its ability to bind the endosomal lipid phosphatidylinositol 3-phosphate (PtdIns(3)P) through its FYVE zinc-finger domain<sup>19,20</sup>. Indeed, PtdIns(3)P binding recruits Hrs, and thereby ESCRT-0, to endosomal membranes<sup>21</sup>.

A third ubiquitin-binding protein, Eps15b, has been found to be associated with ESCRT-0 in human cells. Although its exact function is not known, this protein is important for the function of ESCRT-0 in sorting

<sup>1</sup>Centre for Cancer Biomedicine, Norwegian Radium Hospital, University of Oslo, Oslo, Norway. <sup>2</sup>Department of Biochemistry, Institute for Cancer Research, Rikshospitalet University Hospital, Montebello, 0310 Oslo, Norway.



**Figure 1 | Endocytic internalization and sorting.** Membrane proteins of different kinds are brought inside the cell by endocytosis and translocated to early endosomes, which also receive cargoes from the trans-Golgi network (TGN). Depending on the various proteins that enter the endosome membrane, these cargoes are sorted to distinct destinations. Some cargoes, such as nutrient receptors, are recycled back to the plasma membrane, either directly (step 1a) or indirectly through the recycling endosome (step 1b).

Others, such as receptors for lysosomal enzymes and some protein toxins, are sorted to the TGN (step 2). Ubiquitylated membrane proteins, such as activated growth-factor receptors, are sorted into intraluminal vesicles and eventually end up in the lysosome lumen via MVEs (step 3). In contrast, lysosomal membrane proteins reach their destination by being preferentially sorted to the perimeter membrane of the MVE (step 4). The contents of autophagosomes are also degraded within lysosomes.

epidermal growth factor receptors (EGFRs) for degradation<sup>22</sup>. ESCRT-0 is the least conserved of the ESCRTs and is, for instance, not found in plants<sup>23</sup>. There are likely to be one or more alternative ESCRT-0 proteins that function either in parallel with or instead of Hrs and STAM. TOM and GGA proteins are good candidates for such a function. Like ESCRT-0, these proteins interact with ESCRT-I and contain VHS, ubiquitin-binding and clathrin-binding domains<sup>24,25</sup>, and may associate with PtdIns(3)P-binding proteins that target them to endosome membranes<sup>26</sup>.

### ESCRT-I

The first ESCRT to be characterized was ESCRT-I, initially in yeast and later in mammalian cells<sup>9,27–30</sup>. This complex has a 1:1:1:1 ratio of four subunits: Tsg101 (Vps23 in yeast), Vps28, Vps37 and Mvb12 (refs 29, 30). Crystallographic studies have resulted in a structural model of almost the entire ESCRT-I in yeast<sup>28</sup>. It contains a headpiece (the part that binds ESCRT-II), a rigid 13-nm stalk and an endpiece that contains the ubiquitin- and ESCRT-0-binding UEV domain. The UEV domains of Tsg101 and Vps23 bind to PSAP-like motifs in Hrs/Vps27, and together with additional interactions these motifs contribute to the endosomal recruitment of ESCRT-I. Consequently, in the absence of ESCRT-0, recruitment of ESCRT-I to endosomal membranes is inhibited<sup>10–12</sup>.

### ESCRT-II

ESCRT-II is composed of one Vps36, one Vps22 and two Vps25 subunits<sup>31–33</sup>. The entire core of ESCRT-II consists of eight winged-helix repeats. Vps36 has a split pleckstrin homology domain at its N terminus, which binds 3-phosphorylated phosphoinositides found in the endosome membrane. This domain, known as GLUE (GRAM-like ubiquitin-binding in Eap45), also binds ubiquitin<sup>34–36</sup>. ESCRT-II binds to the ESCRT-I Vps28 carboxy-terminal domain subunit through a helix

immediately C-terminal to the GLUE domain. In yeast Vps36, the GLUE domain contains two inserted NZF zinc-finger domains, one of which binds ubiquitin while the other binds Vps28 (ref. 37). In addition to the lipid-binding properties of the GLUE domain, the first helix of Vps22 also has (less specific) lipid-binding properties that are likely to participate in membrane recruitment of this complex<sup>37</sup>.

### ESCRT-III

This consists of small, highly charged subunits that assemble into higher-order multimers on membranes. The inactive monomeric form of the metastable subunits is maintained by interactions between the autoinhibitory C terminus and the N-terminal part of the subunit<sup>14,38</sup>. During assembly, conformational changes are triggered that relieve the autoinhibition, allowing interactions with other ESCRT-III subunits<sup>14</sup>. This might happen in a directional order, such that one activated ESCRT-III subunit activates the next one. From biochemical studies in yeast, it is now becoming clear that the core complex, which contains the subunits Vps20, Vps32, Vps24 and Vps2, assembles in a highly ordered manner<sup>39</sup>. The subunit that nucleates ESCRT-III assembly on membranes is Vps20, which is N-terminally myristoylated and is thought to interact directly with the endosome membrane. This subunit binds to the Vps25 subunit of ESCRT-II, which functions in recruitment (and possibly in activation) of Vps20 (ref. 33). Because ESCRT-II contains two branches with one Vps25 subunit each, two filaments of ESCRT-III could be nucleated for each copy of ESCRT-II (Fig. 2). Vps20 interacts directly with Vps32, the most abundant of the ESCRT-III subunits in yeast<sup>39</sup>, which triggers the assembly of Vps32 into filamentous oligomers that seem to be capped by Vps24 (refs 39, 40). Finally, Vps2 associates with the Vps24 cap to mediate recruitment of the ATPase Vps4 (ref. 39). The recruitment and activity of Vps4 are regulated by two ESCRT-III-like subcomplexes, Ist1–Did2 and Vta1–Vps60, the exact functions of which are not yet known<sup>41</sup>.



### Capturing the ubiquitin-tagged cargo

ESCRT-0, -I and -II contain ubiquitin-binding subunits that interact directly with ubiquitylated cargo. Epistasis experiments indicate that the ESCRTs function sequentially according to their numerical order<sup>7–9,11</sup>, although biochemical experiments suggest that ESCRT-0, -I and -II might function as a supercomplex.

ESCRT-0 can be considered as a filter that retains ubiquitylated membrane proteins in the endosome membrane (Fig. 3a). The ubiquitin-interacting motifs (UIMs) in ESCRT-0 have low affinity for ubiquitin<sup>42–44</sup>, raising the question of how this complex can function in the efficient sorting of ubiquitylated cargo. One possibility is that simultaneous binding of ESCRT-0 to several ubiquitin moieties in the cargo might strengthen the overall binding. In support of this idea, the ESCRT cargo that has been most extensively studied in mammalian cells, EGFR, is known to become conjugated to several ubiquitin molecules in response to high EGF concentration<sup>45</sup>, either as multiple monoubiquitins or as Lys-63-linked polyubiquitin chains<sup>46</sup>. Moreover, polyubiquitylation favours the interaction of EGFR with ESCRT-0, which contains several UIMs, and also seems to favour degradative sorting of EGFR (ref. 47). It is worth noting that the UIM of Hrs can bind two ubiquitin moieties, and that mutations that abolish this 'double-sided' binding interfere with the sorting function of Hrs<sup>43</sup>. In contrast, yeast Vps27 contains two single-sided UIMs, both of which are important for MVE sorting but not MVE biogenesis<sup>48</sup>, suggesting that there has been evolutionary pressure to maintain the ability of ESCRT-0 to bind multiple UIMs.

Another way for ESCRT-0 to overcome its feeble affinity for ubiquitin could be to increase its concentration locally. This might be accomplished through the ability of Hrs to bind clathrin through a C-terminal clathrin-box motif<sup>17</sup>, which has been shown to increase the concentration of ESCRT-0 in restricted clathrin coats on endosomal membranes, and to enhance the complex's sorting function<sup>42,49,50</sup>. It is tempting to speculate that ESCRT-0 is first recruited to endosome membranes by PtdIns(3)P binding, and is then clustered in microdomains through the ability of clathrin to polymerize (Fig. 3a).

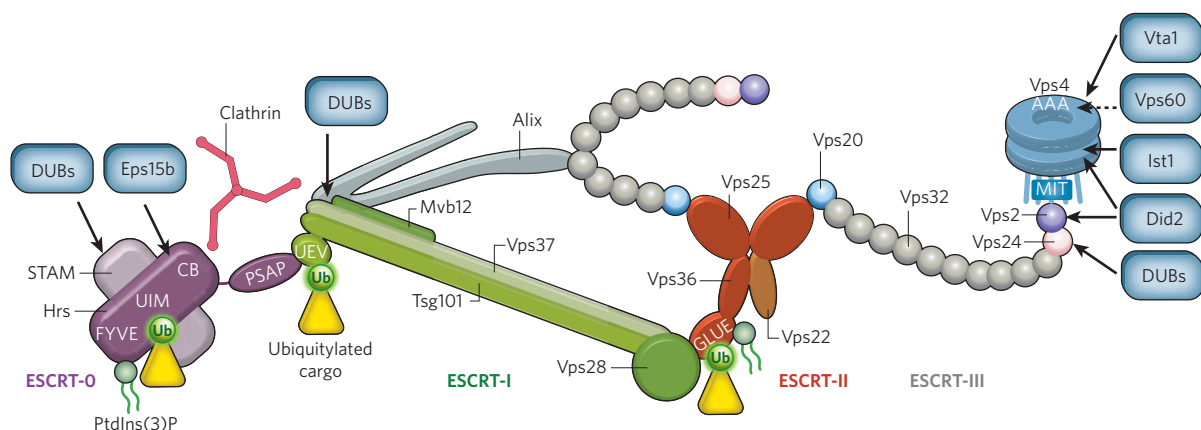
In contrast to ESCRT-0, ESCRT-I and ESCRT-II each have only one ubiquitin-binding domain. They are therefore less suited to the initial capture of cargo and are likely to function when ubiquitin-containing cargo has already been concentrated by ESCRT-0 (Fig. 3a). The ubiquitin-interacting domains of ESCRT-0, -I and -II all bind to the same hydrophobic patch of ubiquitin at Ile 44, so it has been proposed that cargo could be handed over sequentially from one complex to another. However, the long rigid stalk that separates the ubiquitin-binding UEV domain of ESCRT-I from the headpiece that binds the ubiquitin-interacting GLUE domain of ESCRT-II (see Fig. 2) makes cargo exchange within one ESCRT-I/-II pair highly unlikely. This does not exclude the possibility that cargo can be delivered from the UEV domain of one ESCRT-I/-II pair to the GLUE

domain of another pair. Because there is no complete structural model of ESCRT-0, it is not clear whether cargo exchange between individual ESCRT-0 and -I complexes is structurally feasible. Nevertheless, the directional flow of cargo from ESCRT-0 to ESCRT-I and -II is plausible when one considers the sequential recruitment of these complexes. The ubiquitin-binding domains within these complexes have roughly the same affinity for ubiquitin ( $K_D \sim 0.1\text{--}0.3\text{ mM}$ ), raising the question of how sequential cargo binding might be coordinated. One possibility is that cargo binding is regulated by post-translational modifications. The mammalian ESCRT-0 subunits are known to be phosphorylated on specific tyrosine residues<sup>51,52</sup> and to be monoubiquitylated<sup>53</sup>. The latter modification keeps these subunits in an inactive form owing to intramolecular interactions between their UIMs and the appended ubiquitin moiety<sup>54</sup>. Tsg101 in ESCRT-0 is also monoubiquitylated by the ubiquitin ligase Mahogunin, which could switch Tsg101 between active and inactive forms<sup>55</sup>.

### Deforming the endosome membrane

Electron microscopy has shown that yeast that lack ESCRT components have no ILVs in endosomes and vacuoles, and have a peculiar, multi-lamellar endosome<sup>2</sup>. Similar studies of mammalian cells have yielded a more complex picture, as short interfering RNA (siRNA)-mediated depletions of different ESCRT subunits have caused distinct phenotypes. Nevertheless, the consensus view is that ESCRT depletion causes reduced ILV formation in mammalian cells as well, and multilamellar endosomes resembling those of yeast ESCRT mutants have been observed in cells depleted of the ESCRT-I subunit Tsg101 (ref. 56). This begs the question of how ESCRTs have this effect on ILVs.

Given our understanding of canonical vesicle budding mediated by cytosolic coats<sup>57</sup>, it is difficult to understand how vesicles can bud into the lumen of an endosome, which is topologically the opposite. Conspicuously, the elongated yeast ESCRT-I spans about 25 nm, which is approximately the size of ILVs in MVEs in yeast<sup>28</sup>. This led to speculation that, in addition to its role in cargo sorting, ESCRT-I might contribute to membrane deformation<sup>28</sup> (see Fig. 3a, b). Nevertheless, the only experimental evidence so far pertaining to ESCRT-mediated membrane deformation has been obtained with ESCRT-III. High-level over-expression of the ESCRT-III subunit Vps32 in mammalian cells causes protrusion of buds and tubules from the plasma membrane, triggered by spiral filaments formed by Vps32 multimers<sup>58</sup>. Although the plasma membrane is not the normal site of ESCRT assembly, the topology and diameter of these buds and tubules resemble those of ILVs, suggesting that Vps32-mediated plasma-membrane budding reflects key aspects of ILV biogenesis. The finding that plasma-membrane budding is preferentially caused by Vps32, and not by other ESCRT-III subunits, is consistent with the finding that Vps32 is the main constituent of ESCRT-III filaments in yeast<sup>39</sup>.



**Figure 2 | Composition and molecular interactions of the ESCRT machinery.** Interactions between the four ESCRTs are indicated, as are interactions with ubiquitylated cargo and accessory molecules such as PtdIns(3)P, deubiquitylating enzymes (DUBs), Alix and the ATPase Vps4. Mammalian

protein names have been used but the figure is a composite of data obtained from studies of yeast and mammalian ESCRTs. Protein domains are labelled in white. CB, clathrin-box motif. Ub, ubiquitin. Dashed arrow indicates interaction predicted by genetic studies but not yet confirmed biochemically.

An even more detailed insight into the assembly of ESCRT-III filaments comes from *in vitro* reconstitution experiments using recombinant proteins. Such experiments have shown that a truncated Vps2A protein that lacks the C terminus can coassemble with Vps24 into hollow helical tubules about 50 nm across with membrane-interaction sites on the outside of the tubule, whereas Vps4 binds on the inside and disassembles the tubule upon ATP hydrolysis<sup>59</sup>. Both Vps24 and a chimaeric Vps24–Vps2 protein also form hollow tubules; the latter can be disassembled by Vps4 (ref. 60). Single-particle electron-microscopy reconstruction of helical Vps24 filaments shows two-stranded tubules containing both parallel and head-to-head subunit arrangements. Vps32 also self-assembles *in vitro*, but these polymers are too disorganized to be suitable for detailed structural studies. The size, topology and Vps4 sensitivity of the ESCRT-III tubules formed *in vitro* suggest they could aid our understanding of how ESCRT-III assembles and functions on endosomal membranes.

*In vivo* data from yeast suggest more ordered functions for the individual ESCRT-III subunits, and only Vps32 seems to homopolymerize into filaments *in vivo*<sup>39</sup>. Given the structural similarity between the different ESCRT-III subunits, it is possible that high-level expression could enforce polymerization that does not take place *in vivo*, under conditions in which membranes and accessory components are also available. Nevertheless, the *in vitro* assembly experiments provide the first structural insight into how spiral ESCRT-III filaments could form circular arrays and tubular assemblies that are capable of deforming the endosome membrane. The finding that Vps4 acts inside the Vps2A–Vps24 tubules suggests that Vps4 can enter the forming ILVs and depolymerize ESCRT-III from inside<sup>59</sup>, perhaps explaining why ESCRT-III subunits do not enter ILVs<sup>8</sup>.

In mammalian cells, a new player in ILV biogenesis has recently been identified, the PtdIns(3)P-binding protein SNX3 (ref. 61). Depletion of SNX3 prevents ILV biogenesis but not receptor sorting, indicating that it has a specific role in membrane deformation. How this is accomplished is not known, but the fact that depletion of Hrs leads to depletion of SNX3 (ref. 61) suggests that these proteins may form a complex, although this has yet to be demonstrated.

### Membrane abscission from the inside

Membrane abscission during vesicle budding has been well defined for the budding of clathrin-coated endocytic vesicles from the plasma

membrane. The GTPase dynamin is thought to mediate the final membrane severing through a GTPase-driven conformational change<sup>57</sup>. The opposite topology of the ‘inverse’ budding into MVEs requires a distinct machinery specialized for severing thin stalks of membrane filled with cytosol (Fig. 3c). Unlike classical membrane fission, the machinery needed for this type of membrane abscission is recruited from inside the membrane stalk. Equivalent membrane stalks are also severed during cytokinesis and viral budding from the plasma membrane, and accumulating evidence indicates that ESCRT-III and Vps4 represent a highly conserved machinery for this type of membrane fission (Fig. 4). Phylogenetic analyses indicate that ESCRT-III is the most ancient and conserved of the ESCRTs<sup>23</sup>. The other ESCRTs may have evolved later for efficient cargo recognition and perhaps some aspects of membrane deformation. Counterparts to ESCRT-III and Vps4 have been found in archaeobacteria<sup>62,63</sup>, which have no endomembrane system. In these organisms, ESCRT-III and Vps4 function in plasma-membrane abscission to mediate cell division<sup>63</sup>.

### ESCRTs in cytokinesis

Consistent with the idea that the ESCRTs evolved as machinery for cell division, ESCRT-III and Vps4, and to a smaller extent Tsg101, seem to be required for the final membrane abscission step in mammalian cytokinesis<sup>62,64,65</sup>. These proteins are recruited to the midbody during cytokinesis through interactions with Alix, a mammalian homologue of yeast Bro1, and CEP55, a centrosomal coiled-coil homodimer that translocates to the midbody ring during cytokinesis<sup>65,66</sup> (Fig. 4). Alix could engage ESCRT-III to mediate final abscission after the fusion of Golgi- and endosome-derived vesicles has served to narrow the midbody stalk to a thickness of less than 100 nm (ref. 66). Alternatively, ESCRT-I might recruit ESCRT-III through a direct interaction between Vps28 and Vps20, as shown with the corresponding yeast proteins<sup>67</sup>.

### ESCRTs in viral budding

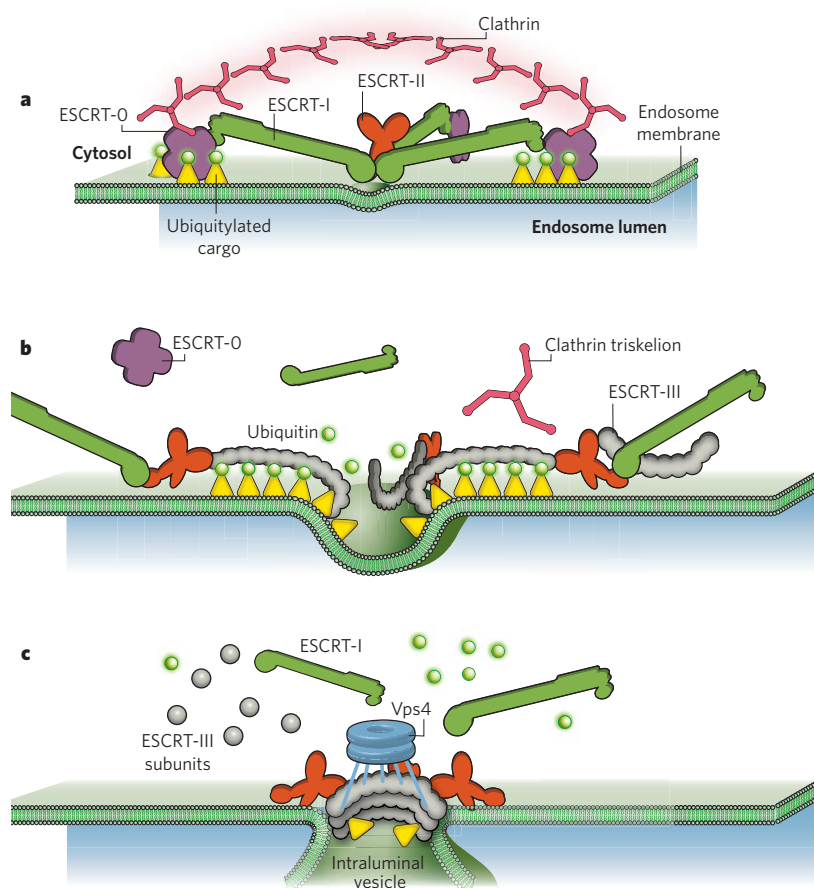
A related function for ESCRTs, Vps4 and Alix is found in the budding of viruses from the plasma membrane of infected cells. The ‘late’ domain of the structural proteins of various viruses contains three alternative sequences, PPXY, P(S/T)AP and LYPXL, which function as recruitment sites for HECT-domain ubiquitin ligases, Tsg101 and Alix, respectively<sup>68,69</sup>. Deletion of these sites causes the failure of membrane

**Table 1 | ESCRT subunits**

Complex	Yeast protein	Metazoan protein	Mammalian synonym	Ubiquitin-binding domain	Selected interacting proteins (metazoan)
ESCRT-0	Vps27	Hrs	NA	UIM	Clathrin, Eps15b
	Hse1	STAM1, 2	NA	UIM (VHS)	AMSH, UBPY
ESCRT-I	Vps23	Tsg101	NA	UEV	NA
	Vps28	Vps28	NA	NA	NA
	Vps37	Vps37A, B, C, D	NA	NA	NA
	Mvb12	Mvb12A, B	NA	NA	NA
ESCRT-II	Vps22	Vps22	EAP30	NA	NA
	Vps25	Vps25	EAP20	NA	NA
	Vps36	Vps36	EAP45	GLUE	NA
ESCRT-III	Vps2	Vps2A,B	CHMP2A,B	NA	NA
	Vps20	Vps20	CHMP6	NA	NA
	Vps24	Vps24	CHMP3	NA	AMSH, UBPY
	Vps32 (Snf7)	Vps32A, B, C	CHMP4A,B,C	NA	NA
Vps4	Vps4	Vps4A, B	SKD1A,B	NA	NA
	Ist1	Ist1	NA	NA	NA
	Did2 (Vps46)	Vps46A, B	CHMP1A,B	NA	NA
	Vta1	Vta1	LIP5	NA	NA
	Vps60 (Mos10)	Vps60	CHMP5	NA	NA
Other	Bro1 (Vps31)	Alix, HD-PTP	NA	NA	AMSH, UBPY

NA, not applicable. In an attempt to simplify, we have mainly used the Vps names of ESCRT subunits, although alternative names exist, especially in mammalian cells, as indicated. Certain mammalian proteins, such as Hrs, Tsg101 and Alix, have such well-established names that their Vps names have been omitted. Note that the assignment of accessory proteins to Vps4 rather than to ESCRT-III is arbitrary in the light of our incomplete understanding of their functions.





**Figure 3 | The ESCRT machinery in endosomal sorting of ubiquitylated membrane proteins.** **a**, Cargo sorting into clathrin-coated microdomains. Initial recognition of ubiquitylated cargo is mediated by ESCRT-0, which is concentrated in microdomains through interaction with clathrin (pink). ESCRT-0 also serves to recruit ESCRT-I. The elongated ESCRT-I recruits ESCRT-II and may contribute to membrane involution. **b**, Membrane deformation. ESCRT-III complexes are recruited by binding to the two Vps25 subunits of ESCRT-II and form spiral-shaped filaments that gate cargo into invaginations that ESCRT-III filaments cause. During this process, cargo is deubiquitylated by DUBs that are recruited by ESCRT-III, but the diffusion of cargo itself is strictly limited by the ESCRT-III filaments. **c**, Membrane abscission. As ESCRT-III filaments assemble into circular arrays, the membrane continues to invaginate. Vps4 enters the invagination to disassemble ESCRT-III filaments, ensuring that its subunits are recycled and that the filaments assemble only at the neck of the forming intraluminal vesicle. For simplicity, the cytosolic part of the transmembrane cargo (yellow) has been omitted in the figure.

abscission and the accumulation of viral buds at the plasma membrane. Although the exact role of ubiquitin in viral budding is unknown, it could act to recruit ESCRT-I through Tsg101 binding. ESCRT-I might mediate ESCRT-III recruitment directly or via Alix, as described for cytokinesis. When comparing the three types of ESCRT-mediated membrane abscission (Fig. 4), it is apparent that ESCRT-0 recruits the ESCRT machinery in ILV budding, whereas CEP55 and viral structural proteins perform the equivalent roles in cytokinesis and viral budding.

Surprisingly, even though ESCRT-II is essential for MVE biogenesis and sorting in yeast, it is dispensable for cytokinesis and the budding of viruses such as HIV-1 (refs 65, 69, 70) and also for ESCRT-mediated lysosomal targeting of ubiquitylated major histocompatibility complex (MHC)-I molecules in mammalian cells<sup>71</sup>. Perhaps there is a direct interaction between Vps28 in ESCRT-I and Vps20 in ESCRT-III (see above), or a bridge between ESCRT-I and ESCRT-III by Alix<sup>72</sup> (Fig. 4), whose Bro1 domain binds to Vps32 in ESCRT-III and whose C-terminal proline-rich domain binds to the UEV domain of Tsg101 in ESCRT-I (refs 72, 73).

### Mechanism of membrane abscission

It seems, then, that ESCRT-0, -I and -II and Alix have accessory roles in membrane abscission, and that a key function of these proteins is to recruit ESCRT-III to the site of membrane involution. Cytokinesis and viral budding bear many similarities in this respect. But how does ESCRT-III mediate abscission? The only known energy input in the severing process comes from the ability of Vps4 to hydrolyse ATP, although it is not clear how this energy is used in membrane fission. The ability of ESCRT-III subunits to assemble into circular arrays holds a key to membrane abscission. It is possible to imagine a constriction of such arrays through Vps4-mediated removal of individual Vps32 subunits from one end of the spiral polymer at the neck of the invagination. This could ultimately cause sufficient constriction to mediate membrane scission<sup>40</sup>. Alternatively, ESCRT-III mediated clustering of cargo with bulky intraluminal domains could also contribute to both membrane deformation and abscission.

### Deubiquitylation of cargo

Although ubiquitylation serves as the signal for entry into the MVE pathway, deubiquitylating enzymes (DUBs) also have important roles in the sorting process. This is illustrated by the finding that Doa4, the DUB that is recruited by ESCRT-III and Bro1 to deubiquitylate endosomal cargo in yeast, is essential for sorting in the MVE pathway<sup>74</sup>. The most straightforward explanation for the requirement of Doa4 is that it mediates the recycling of ubiquitin, thereby avoiding depletion of the cellular ubiquitin pool<sup>75</sup>. However, this is not the only function of Doa4 in the MVE pathway. It has been proposed that Doa4 could also deubiquitylate components of the ESCRT machinery, relieving them of autoinhibition<sup>76</sup>. In mammalian cells, two structurally unrelated endosomal DUBs, AMSH and UBPY, are recruited not only by ESCRT-III and Alix but also by the ESCRT-0 subunit STAM<sup>18,77,78</sup>. The proline-rich domains of these DUBs bind to the same site within the SH3 domain of STAM to mediate their recruitment early in the sorting process. Presumably, these DUBs serve to deubiquitylate cargo that is not destined for the MVE pathway, thereby functioning antagonistically to the E3 ubiquitin ligases that promote MVE sorting. Hypothetically, they could also reverse the autoinhibitory monoubiquitylations of Hrs and STAM. Interestingly, UBPY and AMSH have differential substrate preferences. Although both deubiquitylate Lys-63-linked polyubiquitin chains, UBPY but not AMSH also uses Lys-48-linked polyubiquitin chains as a substrate<sup>77</sup>. This raises the additional possibility that these DUBs might function in the remodelling of ubiquitin chains attached to the cargo.

### ESCRTs and autophagy

Autophagy, or 'self eating', is the process by which cytoplasm, including aggregating proteins, invading microorganisms and damaged organelles, is sequestered by a double-membraned phagophore to form an autophagosome. When the autophagosome fuses with a lysosome, the sequestered cytoplasmic content is degraded by lysosomal enzymes (Fig. 1). A basal level of autophagy has been shown to have a neuroprotective role, although excess autophagy is harmful<sup>79,80</sup>, and the autophagy

of ubiquitylated cytosolic proteins can be considered to represent a second mechanism by which ubiquitin may signal protein degradation in lysosomes. Although it is unclear whether autophagy degrades protein aggregates directly or is important for preventing the accumulation of such aggregates<sup>81</sup>, there is evidence that a ubiquitin-binding protein, p62, mediates both the formation of large ubiquitin-positive protein aggregates and their subsequent clearance<sup>82</sup>. Because p62 interacts with Atg8, which triggers autophagy<sup>83</sup>, it might scavenge small toxic aggregates into larger ubiquitylated aggregates that are presented to the autophagic machinery, thereby mediating the selective autophagy of such structures.

An unexpected connection between the ESCRT machinery and autophagy has recently become apparent. There is a striking increase in the number of autophagosomes when the ESCRT machinery is inactivated, suggesting that this machinery, directly or indirectly, regulates autophagy<sup>84–86</sup>. In principle, such an accumulation of autophagosomes could be caused by elevated formation of autophagosomes, perhaps in response to cellular stress occurring in the absence of ESCRT function. However, it seems more likely that the accumulation is caused by the inhibition of fusion between autophagosomes and lysosomes, as autolysosomes are rare in ESCRT-depleted cells. This raises the question of whether there is any connection between ESCRTs and the machinery that fuses autophagosomes and lysosomes. Molecular interactions have been identified between ESCRT-0 and -I components and Vps18, a component of the HOPS machinery that controls the fusion of autophagosomes and lysosomes (and endosomes and lysosomes)<sup>87</sup>, but so far the functional significance of these interactions is not known. Furthermore, the final closure of the autophagosome (Fig. 1) principally involves the abscission of a thin membrane stalk filled with cytosol — exactly what the ESCRT machinery is designed for (Fig. 4). This raises the possibility that autophagosomes in ESCRT-depleted cells fail to fuse with lysosomes because of inefficient closure,

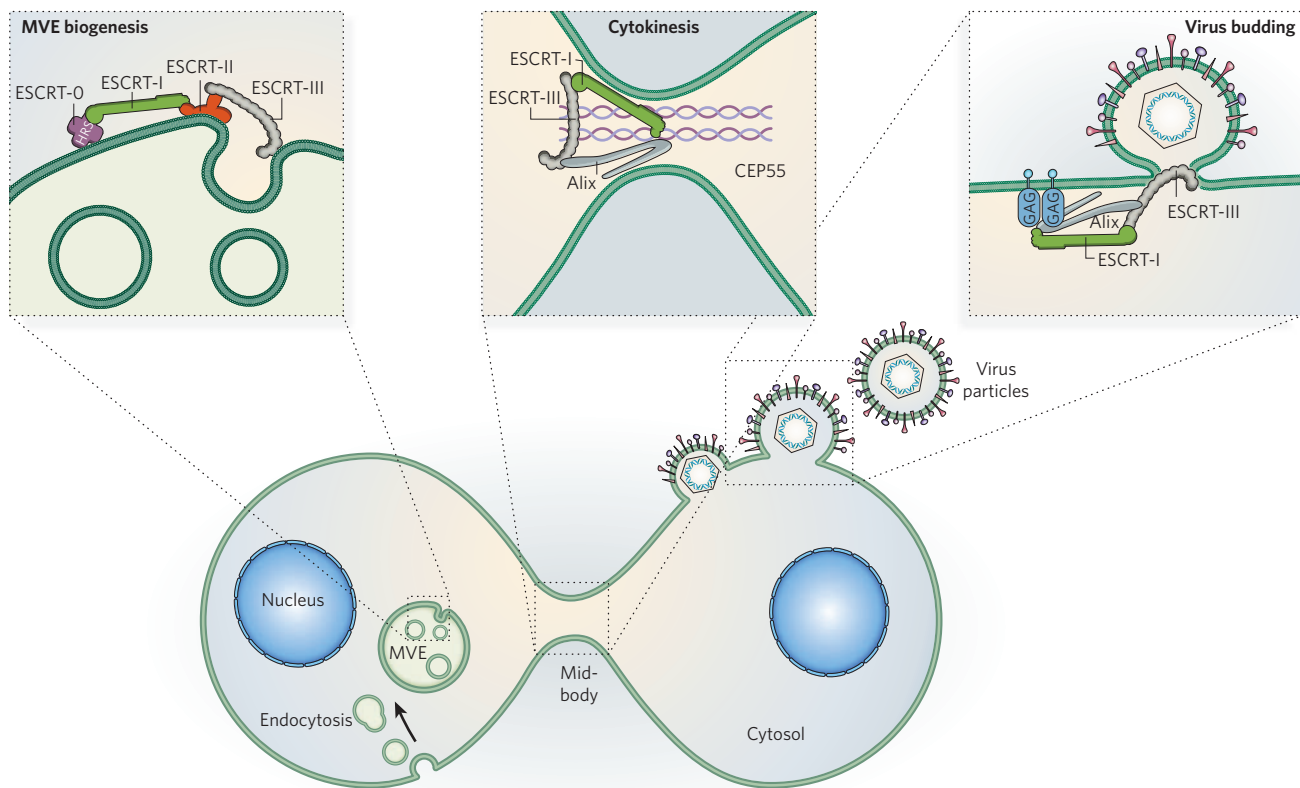
a phenomenon that is difficult to detect using standard fluorescence and electron-microscopic methods.

### ESCRTs fight disease

In addition to their role in viral budding, ESCRT proteins also protect against disorders such as neurodegeneration, cancer and bacterial infections. Individuals with missense mutations in the ESCRT-III subunit Vps2B develop neurodegenerative diseases known as amyotrophic lateral sclerosis and frontotemporal dementia<sup>88,89</sup>. These diseases are characterized by the progressive neuronal accumulation of ubiquitin-positive protein aggregates, perhaps resulting from decreased autophagy<sup>84</sup>. An alternative explanation to the neuronal death caused by Vps2B mutations is the possibility that impaired downregulation of neurotrophin receptors might cause sustained signalling from endosomes that could induce apoptosis<sup>81</sup>.

Sustained receptor signalling is also a key event in carcinogenesis, and Tsg101, Vps37A and Did2 have been identified as putative mammalian tumour suppressors<sup>90–92</sup>. Conditional knockout of Tsg101 in mice has failed to reveal any evidence for its tumour-suppressor activity<sup>93</sup>, but in the fruitfly *Drosophila melanogaster* ESCRT-I and -II (but not ESCRT-0) behave as tumour suppressors<sup>94–96</sup>. Larval tissues expressing clones of ESCRT-I or -II mutant cells form tumours that are largely attributable to the cell non-autonomous stimulation of proliferation caused by excessive cytokine production by the mutant cells. This is triggered by overactive Notch signalling from endosomes, suggesting that the ESCRT machinery is crucial for silencing Notch signalling and thereby for tumour suppression in flies. Whether this is also the case in mammals remains to be clarified.

The engulfment of microorganisms by phagocytes, followed by their destruction when the resulting phagosomes fuse with lysosomes, is a crucial function of the innate immune system. However, certain



**Figure 4 | Involvement of the ESCRT machinery in three topologically equivalent types of membrane abscission.** ESCRT-III is a conserved machinery for the abscission of narrow membrane stalks filled with cytosol. In MVE biogenesis (left), ESCRT-III is recruited by ESCRT-0, -I

and -II. In cytokinesis (middle), ESCRT-III is recruited by the centrosome/midbody protein CEP55 and Alix (and, to a lesser extent, by ESCRT-I). In HIV budding (right), ESCRT-III is recruited by the viral GAG protein and ESCRT-I (and, to a lesser extent, by Alix).



microorganisms, including the mycobacteria that cause tuberculosis, have evolved strategies to subvert the phagolysosomal destruction system, allowing them to thrive and replicate intracellularly. ESCRT components have been identified in a screen of *Drosophila* for factors that restrict the intracellular growth of mycobacteria, and their importance has been verified in mammalian cells<sup>97,98</sup>. There is no evidence that MVE biogenesis as such is important for phagolysosomal function, so the importance of ESCRTs in mycobacterial resistance is likely to have other causes. One reason could be the involvement of ESCRTs in autophagy, a mechanism shown to mediate the killing of mycobacteria<sup>99</sup>. Another possibility is that ESCRTs could mediate the fusion of phagosomes with lysosomes by a mechanism similar to the fusion of autophagosomes with lysosomes.

## Looking ahead

The ESCRTs are derived from a device for cell division in lower unicellular organisms, and have subsequently evolved to recognize ubiquitylated membrane proteins in endosomes, and, following their deubiquitylation, to mediate their translocation into ILVs. It is possible that ESCRT subunits in higher organisms may also have evolved additional functions not related to membrane trafficking.

Advances in structural biology have provided a good picture of the architecture and interaction surfaces of the ESCRTs (Fig. 2), but it remains a major challenge to understand exactly how ESCRT-III assembles on membranes and how this drives inward vesiculation and ILV biogenesis<sup>40</sup>. We also need to understand how ESCRT-0, -I and -II interact with ubiquitylated cargo, and how the cargo is shepherded from its initial contact with ESCRT-0 to the site of endosome invagination. Even though direct cargo transfer between individual ESCRT-I and ESCRT-II complexes is highly unlikely, some means of cargo transfer is required to explain the directional flow of cargo to the budding site. Such delivery could occur in the context of higher-order complexes and might gain directionality by regulated phosphorylation or the ubiquitylation of the ubiquitin-binding subunits. For a thorough understanding of ESCRT-mediated cargo sorting and MVE biogenesis, it might be necessary to reconstitute MVE biogenesis *in vitro* using lipid membranes, purified ESCRT components and ubiquitylated cargo molecules. Experiments of this sort are moving closer as our knowledge of the ESCRTs and their accessory components accumulates. The first *in vitro* reconstitution of ESCRT-III-mediated ILV budding and scission has been published recently, and shows that Vps20, Vps32 and Vps24 are sufficient for detachment of ILVs whereas Vps2 recruits Vps4 to allow recycling of these components for additional rounds of budding<sup>100</sup>. The next step, to reconstitute ESCRT-mediated protein sorting *in vitro*, will be even more challenging, although hopefully achievable in the near future. ■

- Gruenberg, J. & Stenmark, H. The biogenesis of multivesicular endosomes. *Nature Rev. Mol. Cell Biol.* **5**, 317–323 (2004).
- Katzmann, D. J., Odorizzi, G. & Emr, S. D. Receptor downregulation and multivesicular-body sorting. *Nature Rev. Mol. Cell Biol.* **3**, 893–905 (2002).
- d'Azzo, A., Bongiovanni, A. & Nastasi, T. E3 ubiquitin ligases as regulators of membrane protein trafficking and degradation. *Traffic* **6**, 429–441 (2005).
- Hicke, L. & Dunn, R. Regulation of membrane protein transport by ubiquitin and ubiquitin-binding proteins. *Annu. Rev. Cell Dev. Biol.* **19**, 141–172 (2003).
- Hicke, L. & Riezman, H. Ubiquitination of a yeast plasma membrane receptor signals its ligand-stimulated endocytosis. *Cell* **84**, 277–287 (1996).
- Galan, J. M., Moreau, V., Andre, B., Volland, C. & Hagenauer-Tsapis, R. Ubiquitination mediated by the Npi1p/Rsp5p ubiquitin-protein ligase is required for endocytosis of the yeast uracil permease. *J. Biol. Chem.* **271**, 10946–10952 (1996).
- Babst, M., Katzmann, D. J., Snyder, W. B., Wendland, B. & Emr, S. D. Endosome-associated complex, ESCRT-II, recruits transport machinery for protein sorting at the multivesicular body. *Dev. Cell* **3**, 283–289 (2002).
- This paper describes the first biochemical and functional characterization of ESCRT-II.
- Babst, M., Katzmann, D. J., Estepa-Sabal, E. J., Meerloo, T. & Emr, S. D. Escrt-III: an endosome-associated heterooligomeric protein complex required for mvb sorting. *Dev. Cell* **3**, 271–282 (2002).
- This is the first biochemical and functional characterization of ESCRT-III, and identification of Vps20–Vps32 and Vps24–Vps2 as subcomplexes of ESCRT-III.
- Katzmann, D. J., Babst, M. & Emr, S. D. Ubiquitin-dependent sorting into the multivesicular body pathway requires the function of a conserved endosomal protein sorting complex, ESCRT-I. *Cell* **106**, 145–155 (2001).
- This paper coins the ESCRT name and provides the concept of an endosomal machinery that sorts ubiquitylated cargo into MVEs. It describes the first biochemical and functional

characterization of ESCRT-I.

- Bache, K. G., Brech, A., Mehlum, A. & Stenmark, H. Hrs regulates multivesicular body formation via ESCRT recruitment to endosomes. *J. Cell Biol.* **162**, 435–442 (2003).
- Katzmann, D. J., Stefan, C. J., Babst, M. & Emr, S. D. Vps27 recruits ESCRT machinery to endosomes during MVB sorting. *J. Cell Biol.* **162**, 413–423 (2003).
- Lu, Q., Hope, L. W., Brasch, M., Reinhard, C. & Cohen, S. N. TSG101 interaction with HRS mediates endosomal trafficking and receptor down-regulation. *Proc. Natl Acad. Sci. USA* **100**, 7626–7631 (2003).
- Hurley, J. H. ESCRT complexes and the biogenesis of multivesicular bodies. *Curr. Opin. Cell Biol.* **20**, 4–11 (2008).
- Williams, R. L. & Urbe, S. The emerging shape of the ESCRT machinery. *Nature Rev. Mol. Cell Biol.* **8**, 355–368 (2007).
- Prag, G. et al. The Vps27/Hse1 complex is a GAT domain-based scaffold for ubiquitin-dependent sorting. *Dev. Cell* **12**, 973–986 (2007).
- Hofmann, K. & Falquet, L. A ubiquitin-interacting motif conserved in components of the proteasomal and lysosomal protein degradation systems. *Trends Biochem. Sci.* **26**, 347–350 (2001).
- Raiborg, C., Bache, K. G., Mehlum, A., Stang, E. & Stenmark, H. Hrs recruits clathrin to early endosomes. *EMBO J.* **20**, 5008–5021 (2001).
- McCullough, J. et al. Activation of the endosome-associated ubiquitin isopeptidase AMSH by STAM, a component of the multivesicular body-sorting machinery. *Curr. Biol.* **16**, 160–165 (2006).
- Gaullier, J.-M. et al. FYVE fingers bind PtdIns(3)P. *Nature* **394**, 432–433 (1998).
- Burd, C. G. & Emr, S. D. Phosphatidylinositol(3)-phosphate signaling mediated by specific binding to RING FYVE domain. *Mol. Cell* **2**, 157–162 (1998).
- Raiborg, C. et al. FYVE and coiled-coil domains determine the specific localisation of Hrs to early endosomes. *J. Cell Sci.* **114**, 2255–2263 (2001).
- Roxrud, I., Raiborg, C., Pedersen, N. M., Stang, E. & Stenmark, H. An endosomally localized isoform of Eps15 interacts with Hrs to mediate degradation of epidermal growth factor receptor. *J. Cell Biol.* **180**, 1205–1218 (2008).
- Leung, K. F., Dacks, J. B. & Field, M. C. Evolution of the multivesicular body ESCRT machinery; retention across the eukaryotic lineage. *Traffic* **9**, 1698–1716 (2008).
- Puertollano, R. & Bonifacio, J. S. Interactions of GGA3 with the ubiquitin sorting machinery. *Nature Cell Biol.* **6**, 244–251 (2004).
- Puertollano, R. Interactions of TOM1L1 with the multivesicular body sorting machinery. *J. Biol. Chem.* **280**, 9258–9264 (2005).
- Seet, L. F., Liu, N., Hanson, B. J. & Hong, W. Endofin recruits TOM1 to endosomes. *J. Biol. Chem.* **279**, 4670–4679 (2004).
- Stuchell, M. D. et al. The human endosomal sorting complex required for transport (ESCRT-I) and its role in HIV-1 budding. *J. Biol. Chem.* **279**, 36059–36071 (2004).
- Kostelansky, M. S. et al. Molecular architecture and functional model of the complete yeast ESCRT-I heterotetramer. *Cell* **129**, 485–498 (2007).
- By combining data from several crystallographic studies, this paper describes a model of almost the whole ESCRT-I from yeast.
- Morita, E. et al. Identification of human MVB12 proteins as ESCRT-I subunits that function in HIV budding. *Cell Host Microbe* **2**, 41–53 (2007).
- Chu, T., Sun, J., Saksena, S. & Emr, S. D. New component of ESCRT-I regulates endosomal sorting complex assembly. *J. Cell Biol.* **175**, 815–823 (2006).
- Im, Y. J. & Hurley, J. H. Integrated structural model and membrane targeting mechanism of the human ESCRT-II complex. *Dev. Cell* **14**, 902–913 (2008).
- Hierro, A. et al. Structure of the ESCRT-II endosomal trafficking complex. *Nature* **431**, 221–225 (2004).
- This is the first determination of the core of yeast ESCRT-II, showing that the core consists of eight winged-helix domains and that ESCRT-II consists of one Vps22, one Vps36 and two copies of Vps25.
- Teo, H., Perisic, O., Gonzalez, B. & Williams, R. L. ESCRT-II, an endosome-associated complex required for protein sorting: crystal structure and interactions with ESCRT-III and membranes. *Dev. Cell* **7**, 559–569 (2004).
- Slagsvold, T. et al. Eap45 in mammalian ESCRT-II binds ubiquitin via a phosphoinositide-interacting GLUE domain. *J. Biol. Chem.* **280**, 19600–19606 (2005).
- Hirano, S. et al. Structural basis of ubiquitin recognition by mammalian Eap45 GLUE domain. *Nature Struct. Mol. Biol.* **13**, 1031–1032 (2006).
- Alam, S. L. et al. Structural basis for ubiquitin recognition by the human ESCRT-II EAP45 GLUE domain. *Nature Struct. Mol. Biol.* **13**, 1029–1030 (2006).
- Teo, H. et al. ESCRT-I core and ESCRT-II GLUE domain structures reveal central role for GLUE domain in linking to ESCRT-I and membranes. *Cell* **125**, 99–111 (2006).
- This paper shows that the GLUE domain of yeast Vps36 binds Vps28 in ESCRT-I and ubiquitin via yeast-specific NZF zinc fingers, and that another face of the GLUE domain binds the endosomal lipid PtdIns(3)P.
- Lata, S. et al. Structural basis for autoinhibition of ESCRT-III CHMP3. *J. Mol. Biol.* **378**, 816–825 (2008).
- Teis, D., Saksena, S. & Emr, S. D. Ordered assembly of the ESCRT-III complex on endosomes is required to sequester cargo during MVB formation. *Dev. Cell* **15**, 578–589 (2008).
- On the basis of biochemical studies of intact yeast cells, this paper provides a model for the ordered assembly of ESCRT-III on membranes.
- Saksena, S., Wahlman, J., Teis, D., Johnson, A. E. & Emr, S. D. Functional reconstitution of ESCRT-III assembly and disassembly. *Cell* **136**, 97–109 (2009).
- Rue, S. M., Mattei, S., Saksena, S. & Emr, S. D. Novel Ist1-Did2 complex functions at a late step in multivesicular body sorting. *Mol. Biol. Cell* **19**, 475–484 (2008).
- Raiborg, C. et al. Hrs sorts ubiquitinated proteins into clathrin-coated microdomains of early endosomes. *Nature Cell Biol.* **4**, 394–398 (2002).
- This paper shows that Hrs functions in degradative endosomal sorting of ubiquitylated membrane proteins, and that ESCRT-0 is concentrated in restricted microdomains of the endosome membrane.
- Hirano, S. et al. Double-sided ubiquitin binding of Hrs-UIP in endosomal protein sorting. *Nature Struct. Mol. Biol.* **13**, 272–277 (2006).
- Fisher, R. D. et al. Structure and ubiquitin binding of the ubiquitin-interacting motif.

- J. Biol. Chem.* **278**, 28976–28984 (2003).
45. Haglund, K. *et al.* Multiple monoubiquitination of RTKs is sufficient for their endocytosis and degradation. *Nature Cell Biol.* **5**, 461–466 (2003).
  46. Huang, F., Kirkpatrick, D., Jiang, X., Gygi, S. & Sorkin, A. Differential regulation of EGF receptor internalization and degradation by multiubiquitination within the kinase domain. *Mol. Cell* **21**, 737–748 (2006).
  47. Umehayashi, K., Stenmark, H. & Yoshimori, T. Ubc4/5 and c-Cbl continue to ubiquitinate EGF receptor after internalization to facilitate polyubiquitination and degradation. *Mol. Biol. Cell* **19**, 3454–3462 (2008).
  48. Bilodeau, P. S., Urbanowski, J. L., Winistorfer, S. C. & Piper, R. C. The Vps27p Hse1p complex binds ubiquitin and mediates endosomal protein sorting. *Nature Cell Biol.* **4**, 534–539 (2002).
  49. Raiborg, C., Wesche, J., Malerød, L. & Stenmark, H. Flat clathrin coats on endosomes mediate degradative protein sorting by scaffolding Hrs in dynamic microdomains. *J. Cell Sci.* **119**, 2414–2424 (2006).
  50. Sachse, M., Urbe, S., Oorschot, V., Strous, G. J. & Klumperman, J. Bilayered clathrin coats on endosomal vacuoles are involved in protein sorting toward lysosomes. *Mol. Biol. Cell* **13**, 1313–1328 (2002).
  51. Komada, M. & Kitamura, N. Growth factor-induced tyrosine phosphorylation of Hrs, a novel 115-kilodalton protein with a structurally conserved putative zinc finger domain. *Mol. Cell. Biol.* **15**, 6213–6221 (1995).
  52. Takeshita, T. *et al.* Cloning of a novel signal-transducing adaptor molecule containing an SH3 domain and ITAM. *Biochem. Biophys. Res. Commun.* **225**, 1035–1039 (1996).
  53. Polo, S. *et al.* A single motif responsible for ubiquitin recognition and monoubiquitination in endocytic proteins. *Nature* **416**, 451–455 (2002).
  54. Hoeller, D. *et al.* Regulation of ubiquitin-binding proteins by monoubiquitination. *Nature Cell Biol.* **8**, 163–169 (2006).
  55. Kim, B. Y., Olzmann, J. A., Barsh, G. S., Chin, L. S. & Li, L. S. Spongiform neurodegeneration-associated E3 ligase mahogunin ubiquitylates TSG101 and regulates endosomal trafficking. *Mol. Biol. Cell* **18**, 1129–1142 (2007).
  56. Doyotte, A., Russell, M. R., Hopkins, C. R. & Woodman, P. G. Depletion of TSG101 forms a mammalian “Class E” compartment: a multicisternal early endosome with multiple sorting defects. *J. Cell Sci.* **118**, 3003–3017 (2005).
  57. Conner, S. D. & Schmid, S. L. Regulated portals of entry into the cell. *Nature* **422**, 37–44 (2003).
  58. Hanson, P. I., Roth, R., Lin, Y. & Heuser, J. E. Plasma membrane deformation by circular arrays of ESCRT-III protein filaments. *J. Cell Biol.* **180**, 389–402 (2008).  
By using deep-etch electron microscopy of the plasma membrane of cells overexpressing ESCRT subunits, this paper provides spectacular images suggesting that membrane deformation is driven by circular arrays of Vps32 multimers.
  59. Lata, S. *et al.* Helical structures of ESCRT-III are disassembled by VPS4. *Science* **321**, 1354–1357 (2008).
  60. Ghazi-Tabatabai, S. *et al.* Structure and disassembly of filaments formed by the ESCRT-III subunit Vps24. *Structure* **16**, 1345–1356 (2008).
  61. Pons, V. *et al.* Hrs and SNX3 functions in sorting and membrane invagination within multivesicular bodies. *PLoS Biol.* **6**, e214 (2008).
  62. Obita, T. *et al.* Structural basis for selective recognition of ESCRT-III by the AAA ATPase Vps4. *Nature* **449**, 735–739 (2007).
  63. Samson, R. Y., Obita, T., Freund, S. M., Williams, R. L. & Bell, S. D. A role for the ESCRT system in cell division in Archaea. *Science* **322**, 1710–1713 (2008).  
This paper shows that ESCRT-III and Vps4 function in cell division in the archaeobacterium *Sulfolobus acidocaldarius*.
  64. Spitzer, C. *et al.* The *Arabidopsis* elch mutant reveals functions of an ESCRT component in cytokinesis. *Development* **133**, 4679–4689 (2006).  
This is the first demonstration that ESCRT proteins are involved in cytokinesis.
  65. Carlton, J. G. & Martin-Serrano, J. Parallels between cytokinesis and retroviral budding: a role for the ESCRT machinery. *Science* **316**, 1908–1912 (2007).  
This paper provides the first demonstration that ESCRT proteins are involved in mammalian cytokinesis and identifies CEP55 as a key recruiter of ESCRTs to the midbody.
  66. Morita, E. *et al.* Human ESCRT and ALIX proteins interact with proteins of the midbody and function in cytokinesis. *EMBO J.* **26**, 4215–4227 (2007).
  67. Pineda-Molina, E. *et al.* The crystal structure of the C-terminal domain of Vps28 reveals a conserved surface required for Vps20 recruitment. *Traffic* **7**, 1007–1016 (2006).
  68. Martin-Serrano, J. The role of ubiquitin in retroviral egress. *Traffic* **8**, 1297–1303 (2007).
  69. Morita, E. & Sundquist, W. I. Retrovirus budding. *Annu. Rev. Cell Dev. Biol.* **20**, 395–425 (2004).
  70. Langelier, C. *et al.* Human ESCRT-II complex and its role in human immunodeficiency virus type 1 release. *J. Virol.* **80**, 9465–9480 (2006).
  71. Bowers, K. *et al.* Degradation of endocytosed epidermal growth factor and virally-ubiquitinated MHC class I is independent of mammalian ESCRTII. *J. Biol. Chem.* **280**, 5094–5105 (2005).
  72. von Schwedler, U. K. *et al.* The protein network of HIV budding. *Cell* **114**, 701–713 (2003).
  73. McCullough, J., Fisher, R. D., Whitby, F. G., Sundquist, W. I. & Hill, C. P. ALIX-CHMP4 interactions in the human ESCRT pathway. *Proc. Natl Acad. Sci. USA* **105**, 7687–7691 (2008).
  74. Amerik, A. Y., Nowak, J., Swaminathan, S. & Hochstrasser, M. The Doa4 deubiquitinating enzyme is functionally linked to the vacuolar protein-sorting and endocytic pathways. *Mol. Biol. Cell* **11**, 3365–3380 (2000).
  75. Swaminathan, S., Amerik, A. Y. & Hochstrasser, M. The Doa4 deubiquitinating enzyme is required for ubiquitin homeostasis in yeast. *Mol. Biol. Cell* **10**, 2583–2594 (1999).
  76. Nikko, E. & Andre, B. Evidence for a direct role of the Doa4 deubiquitinating enzyme in protein sorting into the MVB pathway. *Traffic* **8**, 566–581 (2007).
  77. Row, P. E., Prior, I. A., McCullough, J., Clague, M. J. & Urbe, S. The ubiquitin isopeptidase UBPY regulates endosomal ubiquitin dynamics and is essential for receptor down-regulation. *J. Biol. Chem.* **281**, 12618–12624 (2006).
  78. Mizuno, E., Kobayashi, K., Yamamoto, A., Kitamura, N. & Komada, M. A deubiquitinating enzyme UBPY regulates the level of protein ubiquitination on endosomes. *Traffic* **7**, 1017–1031 (2006).
  79. Komatsu, M. *et al.* Loss of autophagy in the central nervous system causes neurodegeneration in mice. *Nature* **441**, 880–884 (2006).
  80. Ravikumar, B. *et al.* Inhibition of mTOR induces autophagy and reduces toxicity of polyglutamine expansions in fly and mouse models of Huntington disease. *Nature Genet.* **36**, 585–595 (2004).
  81. Nixon, R. A. & Cataldo, A. M. Lysosomal system pathways: genes to neurodegeneration in Alzheimer’s disease. *J. Alzheimers Dis.* **9**, 277–289 (2006).
  82. Bjørkøy, G. *et al.* p62/SQSTM1 forms protein aggregates degraded by autophagy and has a protective effect on huntingtin-induced cell death. *J. Cell Biol.* **171**, 603–614 (2005).
  83. Pankiv, S. *et al.* p62/SQSTM1 binds directly to Atg8/LC3 to facilitate degradation of ubiquitinated protein aggregates by autophagy. *J. Biol. Chem.* **282**, 24131–24145 (2007).
  84. Filimonenko, M. *et al.* Functional multivesicular bodies are required for autophagic clearance of protein aggregates associated with neurodegenerative disease. *J. Cell Biol.* **179**, 485–500 (2007).
  85. Rusten, T. E., Filimonenko, M., Rodahl, L. M., Stenmark, H. & Simonsen, A. ESCRTing autophagic clearance of aggregating proteins. *Autophagy* **4**, 233–236 (2008).
  86. Lee, J. A., Beigneux, A., Ahmad, S. T., Young, S. G. & Gao, F. B. ESCRT-III dysfunction causes autophagosome accumulation and neurodegeneration. *Curr. Biol.* **17**, 1561–1567 (2007).
  87. Kim, B. Y. & Akazawa, C. Endosomal trafficking of EGFR regulated by hVps18 via interaction of MVB sorting machinery. *Biochem. Biophys. Res. Commun.* Advance online publication doi:10.1016/j.bbrc.2007.08.046 (2007).
  88. Parkinson, N. *et al.* ALS phenotypes with mutations in CHMP2B (charged multivesicular body protein 2B). *Neurology* **67**, 1074–1077 (2006).
  89. Skibinski, G. *et al.* Mutations in the endosomal ESCRTIII-complex subunit CHMP2B in frontotemporal dementia. *Nature Genet.* **37**, 806–808 (2005).
  90. Li, L. & Cohen, S. N. Tsg101: a novel tumor susceptibility gene isolated by controlled homozygous functional knockout of allelic loci in mammalian cells. *Cell* **85**, 319–329 (1996).
  91. Xu, Z., Liang, L., Wang, H., Li, T. & Zhao, M. HCRP1, a novel gene that is downregulated in hepatocellular carcinoma, encodes a growth-inhibitory protein. *Biochem. Biophys. Res. Commun.* **311**, 1057–1066 (2003).
  92. Li, J., Belogortseva, N., Porter, D. & Park, M. Chmp1A functions as a novel tumor suppressor gene in human embryonic kidney and ductal pancreatic tumor cells. *Cell Cycle* **7**, 2886–2893 (2008).
  93. Krempler, A., Henry, M. D., Triplett, A. A. & Wagner, K. U. Targeted deletion of the Tsg101 gene results in cell cycle arrest at G1/S and p53-independent cell death. *J. Biol. Chem.* **277**, 43216–43223 (2002).
  94. Vaccari, T. & Bilder, D. The *Drosophila* tumor suppressor vps25 prevents nonautonomous overproliferation by regulating Notch trafficking. *Dev. Cell* **9**, 687–698 (2005).
  95. Thompson, B. J. *et al.* Tumor suppressor properties of the ESCRT-II complex component Vps25 in *Drosophila*. *Dev. Cell* **9**, 711–720 (2005).
  96. Moberg, K. H., Schelble, S., Burdick, S. K. & Hariharan, I. K. Mutations in erupted, the *Drosophila* ortholog of mammalian tumor susceptibility gene 101, elicit non-cell-autonomous overgrowth. *Dev. Cell* **9**, 699–710 (2005).
  97. Philips, J. A., Porto, M. C., Wang, H., Rubin, E. J. & Perrimon, N. ESCRT factors restrict mycobacterial growth. *Proc. Natl Acad. Sci. USA* **105**, 3070–3075 (2008).
  98. Vieira, O. V. *et al.* Acquisition of Hrs, an essential component of phagosomal maturation, is impaired by mycobacteria. *Mol. Cell. Biol.* **24**, 4593–4604 (2004).
  99. Singh, S. B., Davis, A. S., Taylor, G. A. & Deretic, V. Human IRGM induces autophagy to eliminate intracellular mycobacteria. *Science* **313**, 1438–1441 (2006).
  100. Wollert, T., Wunder, C., Lippincott-Schwartz, J. & Hurley, J. H. Membrane scission by the ESCRT-III complex. *Nature* Advance online publication doi:10.1038/nature07836 (2009).  
This paper is the first to use giant unilamellar vesicles and purified proteins to reconstitute ESCRT-III-mediated ILV abscission *in vitro*, and proves that Vps20, Vps32 and Vps24 are sufficient for ILV budding whereas Vps4 acts to recycle these proteins.

**Acknowledgements** C.R. is a postdoctoral fellow of the Norwegian Cancer Society. We also thank the Research Council of Norway, the Novo-Nordisk Foundation and the Hartmann Family Foundation for supporting our research.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Correspondence should be addressed to H.S. ([stenmark@ulrik.uio.no](mailto:stenmark@ulrik.uio.no)).



# The ubiquitylation machinery of the endoplasmic reticulum

Christian Hirsch<sup>1</sup>, Robert Gauss<sup>1</sup>, Sabine C. Horn<sup>1</sup>, Oliver Neuber<sup>1,2</sup> & Thomas Sommer<sup>1</sup>

**As proteins travel through the endoplasmic reticulum (ER), a quality-control system retains newly synthesized polypeptides and supports their maturation. Only properly folded proteins are released to their designated destinations. Proteins that cannot mature are left to accumulate, impairing the function of the ER. To maintain homeostasis, the protein-quality-control system singles out aberrant polypeptides and delivers them to the cytosol, where they are destroyed by the proteasome. The importance of this pathway is evident from the growing list of pathologies associated with quality-control defects in the ER.**

About 20% of the proteins encoded by the human genome are predicted to be secretory proteins<sup>1</sup>. These proteins pass through the endoplasmic reticulum (ER) on the way to their destination in membranes, exocytotic and endocytotic compartments or the cell exterior. But far from being a passive conduit, the ER is home to an array of molecular chaperones that help proteins to fold and guide their maturation. Despite this support, protein biogenesis is prone to errors. Approximately one-third of all newly synthesized proteins are degraded co-translationally or destroyed within minutes of their synthesis<sup>2</sup>, suggesting that these polypeptides fail to attain their native conformation owing to mutations, transcriptional and translational errors, folding defects or the imbalanced manufacture of subunits. Mature proteins can be damaged by environmental conditions, such as high-energy radiation, chemical insults or metabolic by-products. The malfunction or aggregation of defective proteins challenges the homeostasis of the ER and the cell as a whole. As a result, evolution has produced an ER-resident protein-quality-control system that operates on several levels to maintain the integrity of the ER.

When polypeptides are first synthesized, they are protected from degradation by a specific N-linked glycan structure to allow their maturation. Later, potentially misfolded substrates are flagged up by a unique glycan code generated by mannosidases in the ER (Fig. 1). This signal is decoded by a ubiquitin ligase anchored in the ER membrane. Proteins committed for degradation are transported across the ER membrane, ubiquitylated and degraded by the 26S proteasome, by a process referred to as ER-associated degradation (ERAD). Misfolded proteins are not the only substrates to fall victim to this system; it also regulates sterol synthesis by eliminating the pathway's rate-limiting enzyme when sterols are abundantly available<sup>3</sup>.

Initially, biochemical and genetic screens allowed researchers to identify the components that constitute the ERAD system. Now, the field has moved to functional aspects of this system, as exemplified by recent discoveries on the breakdown of misfolded glycoproteins. Here we review how the ER's protein-quality-control system selectively disposes of aberrant proteins without jeopardizing nascent polypeptides. Because the ERAD pathways seem to be conserved from yeast to mammals, we will use the yeast *Saccharomyces cerevisiae* to outline the fundamental processes and extend our findings with examples from mammalian systems. Figure 2 gives an overview of the relevant yeast and mammalian ERAD factors in their respective cellular compartments.

## Glycans and protein folding

The prime function of the ER's protein quality-control system is to retain unfolded polypeptides in the ER. Hsp70-type and glycan-dependent chaperones bind non-native proteins to prevent their export to the Golgi, and act jointly with oxidoreductases to remodel proteins that have incorrect conformations. The canonical view suggests that the export machinery accepts only proteins that have attained their native fold on the basis of exit signals embedded in the primary sequences of the polypeptide<sup>4</sup>. An alternative model suggests that there is a flexible standard for protein export<sup>5</sup> based on the energetics of a protein's fold and a cell-specific folding environment (Fig. 3 and Box 1).

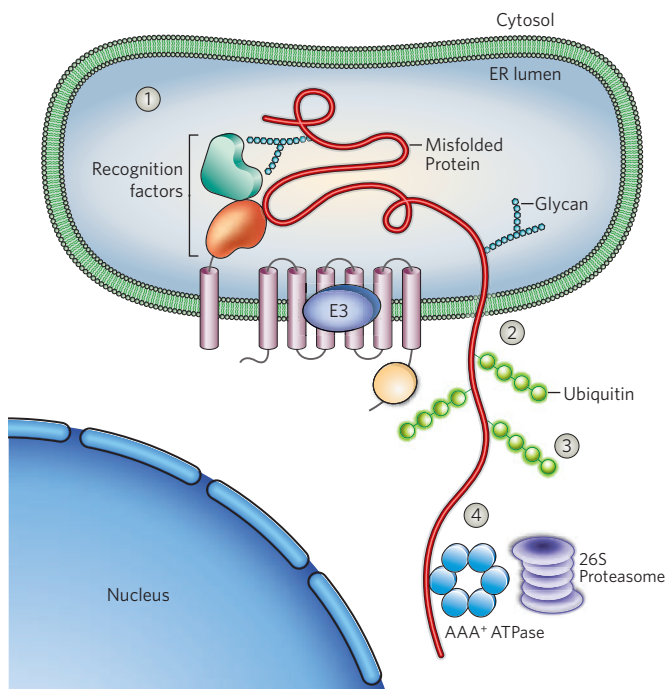
Most polypeptides synthesized in the ER are modified by N-linked oligosaccharides. Oligosaccharyltransferase covalently attaches glucose<sub>3</sub>-mannose<sub>9</sub>-N-acetylglucosamine<sub>2</sub> oligosaccharides to asparagines in Asn-X-Ser/Thr motifs of polypeptides that enter the ER (Fig. 4). Initially, N-glycans render proteins more hydrophilic and direct the co-translational folding of proteins<sup>6</sup>. Moreover, modifications in the oligosaccharide structure provide information about the current folding state of a protein. The two outermost glucose moieties are then immediately removed by  $\alpha$ -glucosidases present in the ER to label polypeptides as in the process of folding (Fig. 4). At this stage, glycan-dependent chaperones associate with the protein. In mammalian cells, calnexin or calreticulin bind to monoglucosylated glycoproteins and promote their maturation<sup>7</sup>. When liberated from these chaperones, ER  $\alpha$ -glucosidase-II trims the residual glucose moiety to prevent a reassociation with calnexin or calreticulin. Glycoproteins that have successfully attained their native structure are allowed to leave the ER<sup>8</sup>.

But what about polypeptides that require more assistance to mature? Being the main folding sensor of the mammalian ER, UDP-glucose:glycoprotein glucosyltransferase (GT, also termed UGGT) recognizes non-native polypeptides and reglucosylates the A branch of oligosaccharides to initiate their reassociation with calnexin or calreticulin for further folding<sup>9</sup> (Fig. 4b). This calnexin-calreticulin cycle serves as a first layer of protein quality control in mammalian cells. However, if a polypeptide has exhausted its chances to mature, further trimming of the oligosaccharide can trigger its destruction.

## Moulding the signal for destruction

Polypeptides that are unable to acquire their native structure must be degraded to prevent fruitless folding attempts and stop the ER from being

<sup>1</sup>Max-Delbrück Center for Molecular Medicine, Robert-Rössle-Strasse 10, 13125 Berlin, Germany. <sup>2</sup>Present address: Institute for Genetics, University of Cologne, Zùlpicher Strasse 47, 50674 Cologne, Germany.



**Figure 1 | Protein degradation at the endoplasmic reticulum (ER).** The degradation of misfolded protein in the ER involves several distinct steps. In step 1, ubiquitin ligases embedded in the ER membrane cooperate with accessory recognition factors to recognize misfolded proteins. Next, in the dislocation phase (step 2), proteins are exported into the cytosol via a so far unidentified channel. At the cytosolic face of the ER, the substrate is ubiquitinated by an E3 ligase (step 3). Finally, the substrate is removed from the membrane by the AAA<sup>+</sup> ATPase Cdc48 and escorted to the 26S proteasome for degradation (step 4).

congested with misfolded polypeptides. The ‘mannose timer’ model proposes that the removal of mannose residues terminates the time a protein is granted to acquire its native conformation<sup>10</sup>. Indeed, members of the glycosylhydrolase-47 (GH47) family in the ER initiate the disposal of misfolded polypeptides. In yeast, the GH47 family encompasses two proteins found in the ER: Mns1 (an  $\alpha$ 1,2-mannosidase) and the mannosidase-like protein Htm1 (homologue to mannosidase 1, also termed Mnl1 for mannosidase-like 1). Mns1 removes the outermost mannose residue from the B branch of N-glycans<sup>11</sup> (Fig. 4). This trimming step affects virtually all glycoproteins in the ER without contributing to glycoprotein maturation or yeast viability<sup>12,13</sup>. Deletion of the *MNS1* gene, however, delays the breakdown of aberrant proteins<sup>13</sup>, suggesting that glycoproteins are protected from degradation until Mns1 converts a mannose<sub>9</sub>-N-acetylglucosamine<sub>2</sub> glycan to a mannose<sub>8</sub>-N-acetylglucosamine<sub>2</sub> glycan.

Because Mns1 processes proteins regardless of their folding status, the mannose<sub>8</sub> structure alone is not sufficient to trigger glycoprotein breakdown; an additional signature must distinguish defective polypeptides from mature proteins. Htm1 was initially proposed to function as a lectin<sup>14,15</sup>, but although direct evidence is still missing, it seems to partake in the removal of the capping  $\alpha$ 1,2-mannose of the C branch from N-glycans that have been processed by Mns1 (ref. 16). The sequential action of Mns1 and Htm1 therefore generates a unique protein-bound mannose<sub>7</sub>-N-acetylglucosamine<sub>2</sub> structure that flags potential ERAD substrates. This idea also explains why deletion of the gene encoding  $\alpha$ 1,3-mannosyltransferase (Alg3) bypasses the need for Htm1 during glycoprotein degradation<sup>13</sup>. The mannose<sub>5</sub>-N-acetylglucosamine<sub>2</sub> structure found on N-glycoproteins of  $\Delta$ alg3 cells is already equipped with the terminal  $\alpha$ 1,6-mannosyl residue generated by Htm1. How Htm1 selects its substrates is not known. It forms a complex with the protein disulphide isomerase Pdi1 (ref. 16), so the action of Htm1 may be restricted to those substrates of Pdi1 that associate with the oxidoreductase for a prolonged time.

In mammalian cells, the GH47 family has diversified into four members found in the ER: ER  $\alpha$ 1,2-mannosidase-I (ERManI) and three mannosidase-like proteins, EDEM1, EDEM2 and EDEM3 (ref. 17). Like its yeast orthologue, mammalian ERManI hydrolyses mannose- $\alpha$ 1,2-mannose bonds, liberating a mannose residue from the B branch<sup>18</sup>. The abundance of ERManI influences the stability of ERAD substrates. Knockdown of ERManI by short interfering RNA (siRNA) impairs degradation of the uncleaved precursor of asialoglycoprotein receptor H2a<sup>19</sup>. Overexpression of ERManI coincides with enhanced de-mannosylation and breakdown of the null Hongkong (NHK) variant of  $\alpha$ 1-antitrypsin<sup>20,21</sup>.

As in yeast, the action of ERManI alone is not sufficient to flag a glycoprotein for degradation. Presumably, ERManI must cooperate with members of the EDEM family to trigger the disposal of aberrant glycoproteins. Although once believed to be lectins, data now suggest that the EDEMs are mannosidases like Htm1. Overexpression of EDEM1 and EDEM3 coincides with mannose trimming of glycans attached to substrates such as NHK, BACE451 and the  $\alpha$ -subunit of the T-cell receptor<sup>22,23</sup>. Moreover, structural modelling of the EDEM proteins unravels striking similarities to ERManI and reveals that all the residues important for the catalytic activity of ERManI are conserved<sup>24</sup>. This suggests that at least EDEM1 and EDEM3 also function as mannosidases. How these enzymes select their substrates remains to be established, however. EDEM1 interacts with calnexin, where it may target glycoproteins that are trapped in the calnexin–calreticulin cycle for degradation<sup>25,26</sup>.

### The ERAD system sorts it out

Misfolded glycoproteins that have been processed by ER mannosidase I and Htm1 must ultimately arrive at a ubiquitin ligase for ubiquitylation. Aberrant substrates that are not glycosylated share this fate, although the events that initiate their degradation are poorly understood. An overview of ERAD substrates is given in Box 2. The heterogeneity of the defects that can occur and the notion that errors may arise in the luminal, transmembrane or cytosolic domain of a protein necessitate distinct E3 ligases, which harness accessory factors to recruit terminally misfolded proteins.

### Different defects require different pathways

In yeast, there are two E3 ligases that target largely discrete substrates. The ligase Doa10 (degradation of Mat- $\alpha$ 2-10) acts on substrates with lesions in their cytosolic domain, whereas substrates of the HRD (HMG-CoA reductase degradation) ligase typically have defects in their transmembrane or luminal domain. To emphasize these differences, an extension of the ERAD term was proposed: ERAD-C degrades proteins with defective cytosolic domains; ERAD-M recognizes lesions in membrane domains; and ERAD-L targets luminal substrates<sup>27,28</sup>. All three pathways commence at different branches and converge in the cytosol. Proteins with multiple defects are preferentially processed by ERAD-C, which seems to operate faster than the other pathways<sup>28</sup>. It should be noted, however, that this terminology is based on a limited set of substrates in yeast and may not be generally applicable.

### Keynotes from ERAD in yeast

The HRD ligase is a membrane-anchored complex comprising at least five distinct subunits. Hrd1/Der3 features six transmembrane domains and a cytosolic RING-finger domain that ubiquitylates target proteins of the ligase<sup>29,30</sup>. Hrd1 interacts with Usa1 (U1-Snp1 associating-1)<sup>27</sup>, the adapter for Der1 (degradation in the ER-1)<sup>31</sup>. The luminal domain of the HRD complex is composed of two proteins: Hrd3 and Yos9 (yeast OS-9 homologue)<sup>27,32–34</sup>. The latter is equipped with a mannose-6 phosphate receptor homology (MRH) domain that recognizes N-glycans containing a terminal  $\alpha$ 1,6-linked mannose<sup>35</sup>, a structure generated with the help of Htm1 (ref. 16). Mutations in the MRH domain abolish binding of N-glycans and thus the breakdown of CPY\* (CPY with a mutation of glycine to arginine at position 255) and other glycosylated polypeptides<sup>35</sup>, suggesting that Yos9 links the recognition of misfolded glycoproteins to the ubiquitin–proteasome system. However, the



deletion of Yos9 or removal of the glycosylation sites in CPY\* does not disrupt the interaction between substrates and the HRD ligase<sup>33,34</sup>. So what is the function of Yos9 or the glycans on CPY\* if they are not required for substrate binding? The MRH domain of Yos9 might scan client proteins for the presence of a mannose<sub>7</sub>-N-acetylglucosamine<sub>2</sub> glycan without engaging in a stable interaction. Instead, other components of the complex may adhere to misfolded polypeptides. Indeed, Hrd3 binds aberrant proteins and has a preference for hydrophobic polypeptides<sup>34</sup>. Also associated with Yos9 is the Hsp70 chaperone Kar2 (karyogamy 2)<sup>33</sup>, suggesting that the luminal domain of the HRD ligase could interact with misfolded proteins by multiple modes.

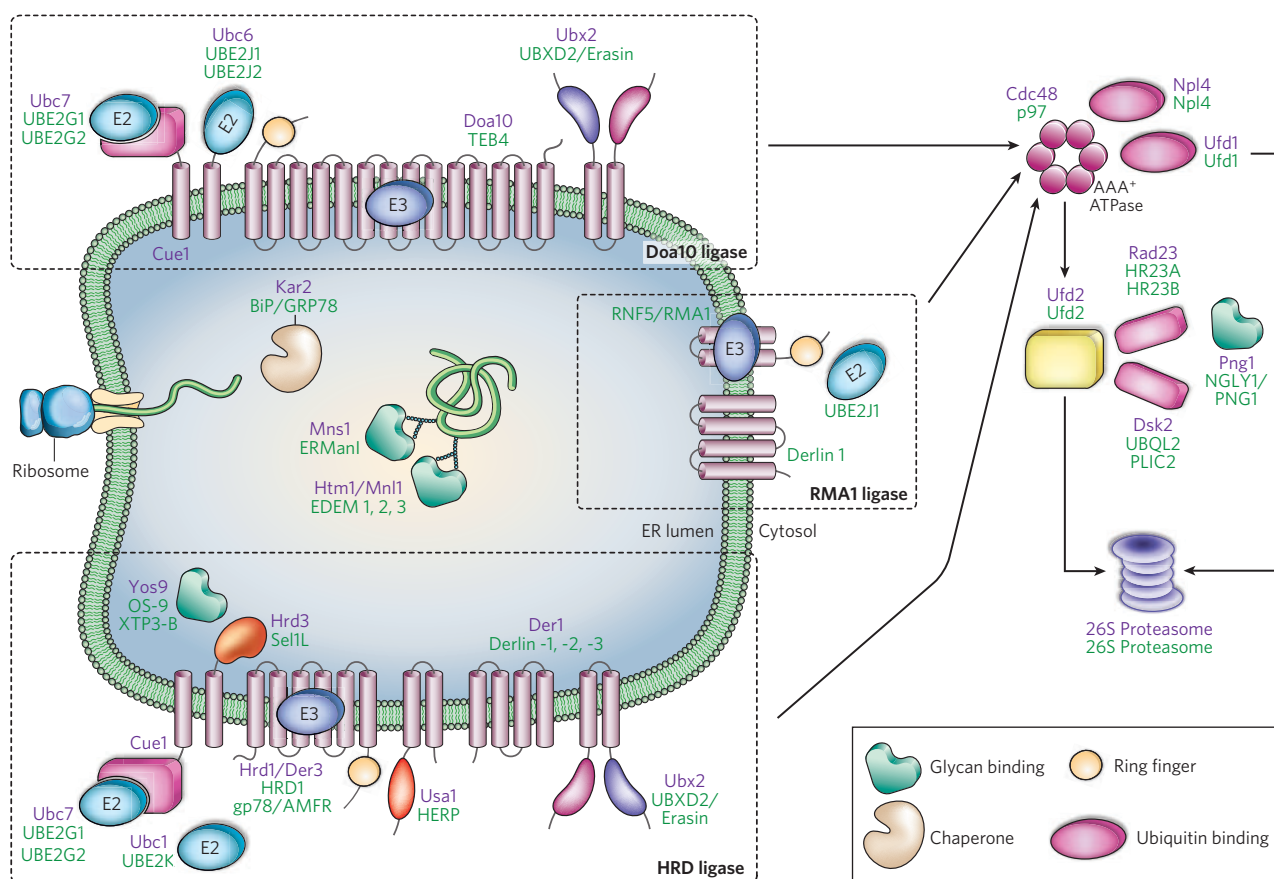
Although speculative, the following model may explain how unwanted polypeptides are selected by the HRD complex for clearance from the ER. Escorted by Kar2, non-native polypeptides arrive at the HRD ligase. These polypeptides can either be terminally misfolded species or proteins that will eventually mature. Next, Hrd3 binds the potential substrate, perhaps mediated by hydrophobic patches within the substrate. Meanwhile, Yos9 inspects client proteins for the presence of a terminal  $\alpha$ 1,6-linked mannose. Only substrates that are concurrently recognized by Yos9 and Hrd3 will be ubiquitinated by Hrd1, with glycoproteins that bear mannose<sub>8-9</sub> glycans being released for further refolding.

Overexpression of Hrd1 in the absence of Yos9 and Hrd3 results in the promiscuous degradation of ER-resident proteins<sup>33</sup>, confirming that Hrd3 and Yos9 may act as gatekeepers to ensure that only legitimate substrates arrive in the cytosol for ubiquitylation. To this end, Hrd1 acts jointly with the soluble E2 enzyme Ubc1 (ubiquitin-conjugating enzyme 1) and with Ubc7, which is recruited to the ER membrane through interaction with Cue1 (coupling of ubiquitin conjugation to ER degradation 1)<sup>29,36</sup>.

The second E3 ligase present in the ER membrane in yeast is Doa10. This protein was first isolated in a screen for factors that degrade the soluble transcriptional repressor Mata2 (ref. 37). With the identification of several membrane-anchored substrates, Doa10 was also linked to the ERAD pathway. Doa10 features an amino-terminal RING-finger and 14 transmembrane domains<sup>38</sup>. Despite this intricate membrane topology, the subunit composition of the complex is relatively simple: the only identified cofactors of Doa10 are the E2 enzymes Ubc6 and Ubc7. The latter is recruited to the ER membrane by Cue1 (ref. 36).

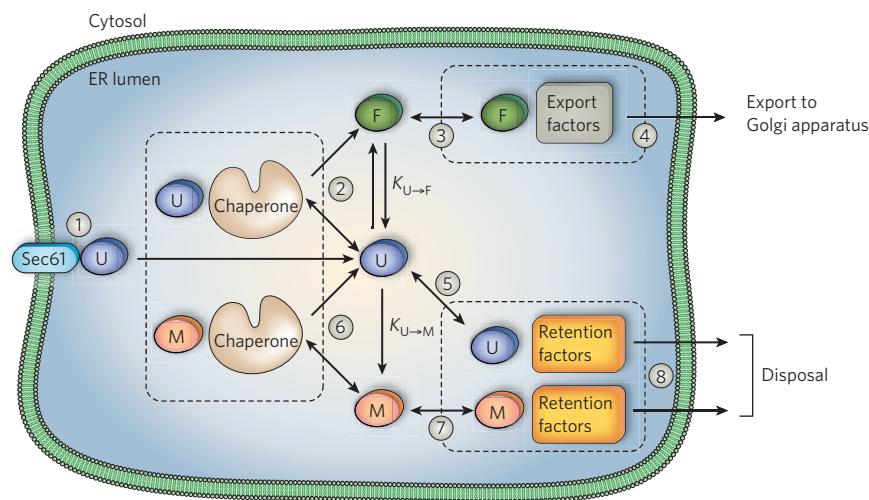
### Mammalian ligases

Perhaps not unexpectedly, many more ligases have been characterized that participate in ERAD in mammals than in yeast. Two mammalian proteins with homology to yeast Hrd1 are known: HRD1 (or Synoviolin) and gp78 (also termed AMFR, for tumour autocrine motility factor receptor)<sup>39-41</sup>. HRD1 is part of a complex related to the yeast HRD ligase as it involves the human Hrd3 orthologue SEL1L (suppressor of Lin12-like protein)<sup>42</sup>, HERP (Usa1 in yeast)<sup>43,44</sup>, Derlin 1-3 (a family of Der1 representatives)<sup>45,46</sup> and OS-9 (amplified in osteosarcoma, the human orthologue of Yos9)<sup>47,48</sup>. Mammalian cells express two forms of OS-9: OS-9.1 and a splice variant, OS-9.2, which lacks 55 amino-acid residues. Both proteins interact with SEL1L and bind the misfolded NHK variant of  $\alpha$ 1-antitrypsin<sup>47,48</sup>. Consistent with reports on their yeast counterpart Yos9, binding does not require N-linked glycans, as the interaction persists when a non-glycosylated variant of NHK (NHK-QQQ) is used as substrate<sup>47</sup>. However, a lack of glycans does not result in the degradation of the client protein. Reduction of OS-9 levels by siRNA impairs NHK degradation and increases the amount of



**Figure 2 | Mechanism of protein degradation in yeast and mammalian cells.** Molecular chaperones and proteins of the glycosylhydrolase-47 family (Mns1 and Htm1) detect misfolded polypeptides and direct them to membrane-bound ligases (Doa10, RMA1, HRD). After dislocation to the cytosolic side of the ER membrane, substrates are ubiquitinated by an E3 ligase. All ligase complexes comprise a central, catalytic RING-finger

protein (E3), ubiquitin-conjugating enzymes (E2) and additional factors. The AAA<sup>+</sup> ATPase Cdc48 releases ubiquitylated molecules from the ER membrane. The adaptor proteins Rad23 and Dsk2 escort the ubiquitylated molecules to the 26S proteasome for degradation; concurrently Png1 deglycosylates glycoproteins through its association with Rad23. Yeast proteins are labelled in purple, their mammalian counterparts in green.



**Figure 3 | Protein homeostasis in the ER.** Proteins are translocated into the ER by Sec61 (step 1). Unfolded proteins (U) can engage chaperones, which help them fold into their native conformation (F, step 2). Export factors select correctly folded proteins (step 3) and transport them to the Golgi (step 4). Retention factors prevent the exit of unfolded proteins (step 5). Chaperones try to remodel any misfolded proteins (M) into a folding-competent, unfolded state (step 6). Retention factors keep misfolded proteins in the ER (step 7). Eventually, defective proteins, along with a fraction of folding intermediates (U), are exported to the cytosol for disposal (step 8). The equilibrium constants for the folding ( $K_{U \rightarrow F}$ ) and misfolding ( $K_{U \rightarrow M}$ ) reactions are indicated.

NHK secreted. Overexpression of OS-9 primarily reduces the secretion of NHK without enhancing degradation<sup>47</sup>. Thus OS-9 might perform two tasks: at physiological concentrations, it cooperates with SEL1L to facilitate the disposal of misfolded proteins; when overexpressed, excess OS-9 is unable to contact the HRD ligase and retains non-native proteins in the lumen of the ER.

Another mammalian orthologue of Yos9 is XTP3-B (XTP3-transactivated protein B, or Erlectin)<sup>47,49</sup>. This protein has two MRH domains and interacts with SEL1L. Like OS-9, XTP3-B binds NHK and its unglycosylated variant. Overexpression of XTP3-B delays the degradation of both substrates, whereas siRNA against XTP3-B has no influence on NHK stability<sup>49</sup>. Suitable substrates for XTP3-B may still await identification; alternatively, XTP3-B may retain immature proteins in the ER and prevent their aggregation, much like overexpressed OS-9.

HRD1 protects cells from stress-induced apoptosis<sup>50</sup>, suggesting a broader role in the ubiquitylation of damaged ER proteins. Among the substrates of the HRD ligase identified are a truncated and misfolded form of Ribophorin termed RI332 (ref. 42), unassembled secretory Igu chains<sup>51</sup> and a non-glycosylated variant of the Igk light chain<sup>52</sup>. In the ER, this Igk light chain exists in a fully oxidized form equipped with two

disulphide bonds and a partly reduced variant that has one disulphide bond<sup>52</sup>. Only the partly reduced molecule interacts with both Derlin-1 and HRP, suggesting that one disulphide bond in the oxidized form of the Igk light chain is reduced, before the light chain is recruited to the ligase<sup>52</sup>. The human cytomegalovirus (HCMV) gene product US11 (unique short region protein 11) also uses components of the HRD ligase as it requires SEL1L and Derlin-1 to target MHC-class-I heavy chains for destruction<sup>42,45,46</sup>.

Gp78 and human HRD1 have more than 50% sequence homology in their transmembrane domains. Unlike Hrd1, which requires Cue1 to associate with Ubc7, gp78 is equipped with a G2BR (UBE2G2-binding region) domain distal to its RING-finger domain that recruits the E2 UBE2G2 (ubiquitin-conjugating enzyme E2G 2, the orthologue of yeast Ubc7) to the ligase<sup>53</sup>. Among the targets of gp78 is the Pi Z variant of  $\alpha 1$ -antitrypsin<sup>54</sup> and orphan subunits of the T-cell receptor, such as TCR $\alpha$ <sup>53</sup> and Cd3 $\delta$ <sup>39</sup>. Furthermore, gp78 ubiquitylates HMG-CoA reductase<sup>55</sup>, implying that the role of the ERAD system in sterol metabolism is broadly conserved. Another gp78 substrate of medical importance is KAI1, a tetraspanning protein that suppresses cancer metastasis<sup>56</sup>. Downregulation of gp78 by siRNA increases KAI1 levels and coincides

#### Box 1 | A flexible standard for protein export

It is commonly believed that the export machinery accepts only properly folded proteins, with aberrant polypeptides being degraded. This suggests that the endoplasmic reticulum (ER) uses a fixed quality-control standard based on the conformation of the wild-type protein. But this idea cannot explain observations such as cell-specific secretion efficiencies of mutant proteins<sup>97,98</sup>. What global principles govern protein quality control in the ER?

In an attempt to understand the basis of tissue-specific amyloid diseases caused by transthyretin (TTR), Sekijima *et al.* demonstrated that the secretion of TTR mutants correlates with a 'folding stability score'<sup>99</sup>. This variable integrates thermodynamic stability, defined as the Gibbs free-energy difference between the elongated polypeptide chain and the folded state, and kinetic parameters, such as the rate of TTR tetramer assembly. Accordingly, discrepancies in secretion efficiency in different tissues seem to be based on the varying activities of chaperones, export machineries and ERAD components that set cell-specific standards for export. For example, a particular protein might be exported in one cell type despite being degraded in another tissue because decreased chaperone activity lowers the 'folding stability score' threshold applied for export.

On the basis of this study, Wiseman *et al.* developed a quantitative model of folding for export (FoldEx)<sup>5</sup> that reduced the complex homeostasis network of the ER to a few basic pathways (translocation, chaperone-independent and chaperone-assisted folding, protein export and ERAD), each considered as an individual unit (Fig. 3). Treated like

enzymes, the function of each pathway was described with Michaelis-Menten kinetics. Combining the different equations, the authors described the export efficiency mathematically depending on the folding equilibrium constant, which is related to the thermodynamic stability. This theoretical approach correctly describes the observation that in a given cell-specific environment the absolute thermodynamic stability of the protein's fold defines its export efficiency. It also explains why even very stable proteins, including multidomain proteins such as CTFR or multi-subunit complexes, can be degraded if they fold slowly. The model reveals that export efficiency depends on the ratio of ERAD and export activities and the rate of misfolding; that is, export efficiency decreases as protein misfolding becomes more likely.

The FoldEx model emphasizes that an 'adaptable standard' determines the export efficiency of each individual protein in a given cell type. This standard is established by the complex interplay of the protein's primary sequence and the local environment of the folding compartment (for example, chaperones and folding enzymes; metabolites; export and ERAD machineries), and can be adjusted to counterbalance conditions of cellular stress. Different proteins will compete for these machineries and, consequently, the export efficiency of a particular protein depends on the proteome that is expressed at a certain time.

The FoldEx model is the first approach to define a theoretical framework that combines the thermodynamic and kinetic parameters of protein folding and the adjustable activities of ER folding, ERAD and transport machineries. It is likely to prove a useful tool for further experiments.



with a reduced metastatic aggressiveness of the tumour cells<sup>56</sup>.

A likely candidate for the mammalian orthologue of Doa10 is TEB4 (ref. 57). Both proteins share the same transmembrane topology<sup>38</sup> and an N-terminal RING-finger domain. Although TEB4 ubiquitylates itself, substrates of this ligase are yet not known<sup>57</sup>.

Apart from these ligases, which share obvious similarities with yeast enzymes, several mammalian ERAD ligases have been characterized that have no equivalent in fungi. The ER protein RMA1 (RING-finger protein with membrane anchor 1) is a ubiquitin ligase conserved from *Arabidopsis* to humans<sup>58</sup>. RMA1 interacts with UBE2J1, and overexpression of the ligase reduces the abundance of the cystic fibrosis transmembrane conductance regulator (CFTR)<sup>58</sup>. RMA1 primarily monitors the folding of the N-terminal domain of CFTR during or soon after translation<sup>58</sup>. Post-translationally, the cytosolic E3 ligase CHIP (carboxyl terminus of Hsp70-interacting protein) relieves RMA1 and monitors the folding of CFTR<sup>58</sup>. CHIP encompasses a U-box that is structurally related to the RING-finger motif and a tetratricopeptide repeat domain that interacts with Hsp70 and Hsp90 (ref. 59). The combination of chaperone binding and ubiquitin–protein ligase activity suggests that CHIP supervises Hsp70-mediated protein folding and triggers the proteasomal degradation of substrates that are irreversibly misfolded.

Another factor that cooperates with RMA1 is BAP31 (B-cell receptor-associated protein 31). This integral membrane protein is involved in the sorting of a diverse set of ER membrane proteins. BAP31 interacts with components of the Sec61 pore, the E3 protein RMA1 and Derlin-1 (ref. 60). Because siRNA-mediated knockdown of BAP31 stabilizes CFTR, BAP31 may chaperone newly synthesized CFTR polypeptides and divert variants that are refractory to folding through RMA1 and Derlin-1 for degradation.

Mutations in the ubiquitin ligase Parkin are associated with juvenile Parkinson's disease<sup>61</sup>. Expression of defective Parkin triggers the accumulation of Pael-R in the ER, followed by a degeneration of dopaminergic neurons, a hallmark of Parkinson's disease. Parkin ubiquitylates Pael-R *in vitro*, albeit with low efficiency. In combination with CHIP, this ubiquitylation is significantly enhanced<sup>62</sup>.

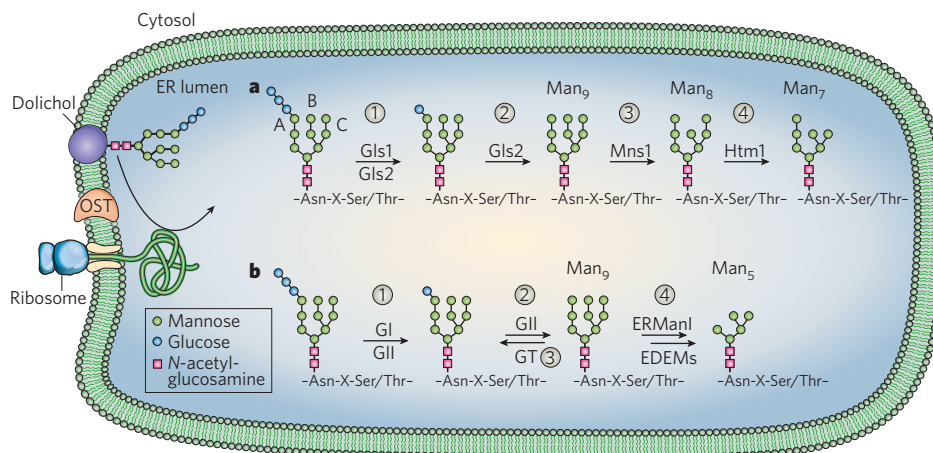
The cytosolic SCF complex is a ubiquitin ligase comprising Skp1, Cul1, Roc1 (also called Rbx1) and an F-Box protein. Among the large set of F-Box proteins that determine the substrate specificity of the SCF

ligase are two neuron-specific factors, Fbs1 and Fbs2 (F-box protein that recognizes sugar chains). Both proteins have a sugar-binding domain that binds to the innermost portion of N-linked glycans<sup>63</sup>. The substrates of the SCF<sup>Fbs1</sup> complex have not yet been characterized, but it may clear the cytosol of glycoproteins that escaped the classical ERAD pathways or leaked from the ER through damage in the lipid bilayer.

### Crossing the membrane

Before misfolded proteins can be ubiquitylated and degraded, they must cross the ER membrane. This export step is referred to as dislocation or retrotranslocation. How proteins are extracted from the ER is not known, but ideas proposed range from leaving via the Sec61 translocator<sup>64,65</sup>, in a reaction that resembles a reversal of the import process, to the formation of lipid droplets that facilitate the escape of polypeptides from the ER<sup>66</sup>.

The HRD ligase links important events on both sides of the ER membrane: substrate recruitment at the luminal side, and protein ubiquitylation at its cytosolic face<sup>33,34,67</sup>. It follows that the conduit for protein export should be in the proximity of the ligase. Perhaps the HRD complex itself discharges aberrant proteins into the cytosol. Associated with the HRD ligase are members of the Derlin family, which are considered prominent candidates for components of the translocation channel. The three proteins of this group show weak homology to yeast Der1 and form hetero-oligomers with each other<sup>45,46</sup>. Derlin-1 interacts with the HCMV protein US11, and MHC class I heavy chains are also present in this complex<sup>45,46</sup>. When the proteasome is inhibited pharmacologically, deglycosylated heavy chains occur in association with Derlin-1 (ref. 45). N-glycanase, the enzyme that deglycosylates class I heavy chains, is restricted to the cytosol<sup>65,68,69</sup>. The association between Derlin-1 and the heavy-chain molecules must therefore occur either during or immediately after dislocation of the substrate. Additional evidence that Derlin proteins may channel aberrant ER proteins to the cytosol comes from the observation that the ATPase associated with various cellular activities, the AAA<sup>+</sup>-ATPase p97 (Cdc48 in yeast), interacts with Derlin-1 and Derlin-2 (refs 70, 71). As discussed below, Cdc48 contributes to the driving force that mobilizes ERAD substrates from the ER membrane. Derlin-1 is not required for the US2-mediated degradation of heavy chains<sup>45</sup>, so it may serve a different function altogether, or alternative exits from the ER must exist.



**Figure 4 | Processing of N-linked glycans in the yeast and mammalian ER.** **a**, In yeast and mammals, oligosaccharyltransferase (OST) transfers preassembled glucose<sub>3</sub>-mannose<sub>9</sub>-N-acetylglucosamine<sub>2</sub> oligosaccharide from a dolichol carrier to asparagine residues in Asn-X-Ser/Thr motifs in the emerging peptide. In yeast, Glc1 and Glc2 (glucosidases 1 and 2) remove the two outermost glucose residues and generate a glucose<sub>1</sub>-mannose<sub>9</sub>-N-acetylglucosamine<sub>2</sub> sugar (step 1). Further processing by Glc2 results in a mannose<sub>9</sub>-N-acetylglucosamine<sub>2</sub> glycan that protects the glycoprotein from disposal (step 2). Mns1 trims the outermost mannose moiety of the B (middle) branch, yielding a mannose<sub>8</sub>-N-acetylglucosamine<sub>2</sub>

oligosaccharide, which indicates a protein that is retained in the ER for a prolonged period of time (step 3). Htm1 then processes the C (right-hand) branch, yielding a mannose<sub>7</sub>-N-acetylglucosamine<sub>2</sub> oligosaccharide (step 4). **b**, In mammalian cells, steps 1 and 2 resemble those in the yeast system, yet removal of the last glucose residue from the A branch by glucosidase-II (GII, step 2) is reversible. Glucosyltransferase (GT) reglucosylates glycoproteins that have not attained their native fold (step 3). Extensive demannosylation by ERManI and probably the ER degradation enhancing EDEM proteins yield a mannose<sub>5</sub>-N-acetylglucosamine<sub>2</sub> glycan that is a signal for disposal (step 4).

## Membrane release and destruction

The movement of aberrant proteins across the ER membrane requires a force to provide directionality. Additionally, substrates must be released from the membrane to become accessible to the proteasome. Both transport and release require energy and may be coupled. The energy is in part required for the ubiquitylation of the substrate, which is a prerequisite for dislocation<sup>72,73</sup>. Another energy-dependent step is mediated by Cdc48 (p97 in mammals)<sup>73–77</sup>. This AAA-ATPase forms homohexamers that associate with the cofactors Ufd1 (ubiquitin fusion degradation 1) and Npl4 (nuclear protein localization 4) to mobilize aberrant ER proteins from the ER membrane. The molecular mechanism of this process is unclear. According to the ‘molecular ratchet’ model<sup>78</sup>, coordinated cycles of ATP hydrolysis induce conformational changes within the complex and propel the membrane release of substrates that attach to the N terminus of p97. In a different model, the guanidyl-rich environment of an arginine-rich ring at the base of the complex forms a ‘denaturation collar’ that unfolds substrates through denaturation<sup>79</sup>. To exert mechanical force on substrates, the Cdc48 complex must be fastened to the ER membrane. Consequently, Cdc48 interacts with the UBX (ubiquitin regulatory X) domain of the membrane-anchored protein Ubx2 by a mechanism that depends on a ubiquitylated substrate<sup>80,81</sup>. The mammalian candidate protein for this function is UBXD2 (or ERASIN)<sup>82</sup>.

When the substrates are polyubiquitylated and extracted from the membrane, they are forwarded to the proteasome by the action of Rad23

(radiation sensitive 23) and Dsk2 (dominant suppressor of Kar1)<sup>83,84</sup>. These homologous adaptor proteins have a C-terminal UBA (ubiquitin-associated) motif that binds polyubiquitin chains linked by a lysine at ubiquitin residue 48 and an N-terminal UBL (ubiquitin-like) domain to guide adaptor and substrate to the proteasome, which recognizes polyubiquitylated substrates through its subunits Rpn10 (regulatory particle non-ATPase 10)<sup>85</sup> and Rpn13 (refs 86, 87). In addition, Dsk2 and Rad23 may shield the polyubiquitin chain from the action of ubiquitin hydrolases. In this canonical pathway, the E3 ligase alone polyubiquitylates substrates. In yeast there is a second route, in which the substrate is first mono- or diubiquitylated by the ligase<sup>84</sup>. Further chain elongation is catalysed by the U-Box protein Ufd2, which interacts with Cdc48 (ref. 84). Cdc48 limits the action of Ufd2 to the attachment of a short oligoubiquitin chain of four to six molecules, which is large enough for recognition by Rad23 and Dsk2 (ref. 84). This incremental increase of the ubiquitin chain may provide the directionality required to target substrates to the proteasome.

Because N-glycanase interacts with Rad23 (ref. 88) and possibly with other components of the ERAD machinery, glycoproteins are liberated from their oligosaccharides when they reach the cytosol. This trimming step is not crucial for degradation<sup>89,90</sup>, but removal of the bulky side groups may facilitate the threading of substrates into the channel that leads to the active-site chamber of the proteasome. Before proteasomal degradation, deubiquitylating enzymes remove the polyubiquitin chain from the substrate and the ubiquitin moieties are recycled<sup>91–93</sup>.

## Box 2 | ERAD substrates

### $\alpha$ 1-Antitrypsin ( $\alpha$ 1-AT)

A soluble, glycosylated serine-protease inhibitor secreted by hepatocytes. The NHK (null Hong Kong) variant harbours a premature stop codon. Although folding-defective, a fraction of the NHK variant escapes the ER quality control and is secreted. NHK-QQQ is an unglycosylated, engineered variant. In the Pi Z variant, a mutation of glutamic acid to lysine at position 342 interferes with folding and prevents secretion. In certain cases, Pi Z accumulates in the ER, leading to liver cirrhosis. The lack of  $\alpha$ 1-AT triggers lung emphysema, owing to increased activity of neutrophil elastase, which disrupts the connective tissue.

### ASGPR H2a

A subunit of the asialoglycoprotein receptor. This receptor clears desialylated glycoproteins from the circulation by receptor-mediated endocytosis.

### BACE457

A pancreatic isoform of the  $\beta$ -secretase (BACE) that does not contribute to the processing of the amyloid precursor protein. Instead, BACE457 is degraded.

### CD3 $\delta$ and TCR $\alpha$

Members of the T-cell receptor, which is assembled in the ER. Orphan subunits are degraded. Likewise, secretory Ig- $\mu$  is degraded in the absence of the light chain.

### CFTR (cystic fibrosis conductance regulator)

A chloride channel expressed by epithelial cells. The  $\Delta$ F508 mutation is the most common cause of cystic fibrosis, an inherited autosomal recessive disease. The  $\Delta$ F508 deletion only marginally impairs the biological activity of the channel, yet maturation of mutant protein is highly ineffective. The lack of osmotically active chloride ions results in the reduced hydration of the mucus produced by epithelial cells. The high viscosity of the mucus impairs clearance of the airways and causes persistent microbial infections, which result in lung fibrosis and failure.

### CPY (carboxypeptidase Y)

A soluble serine protease active in the vacuole of the yeast *Saccharomyces cerevisiae*. CPY carries four N-linked glycans. CPY\* has a mutation of glycine to arginine at position 255. CPY\* that lacks a C-terminal oligosaccharide is not recognized by ERAD.

### HMG-CoA reductase (3-hydroxy-3-methyl-glutaryl-CoA reductase)

A membrane-anchored protein. It is the rate-limiting enzyme of the mevalonate pathway. Its stability is tightly linked to the abundance of sterols. When sterol levels are sufficient, the enzyme is degraded.

### Ig $\kappa$ LC NS

A transport-incompetent immunoglobulin light chain that is degraded by ERAD.

### KAI1 (also known as CD82)

A member of the tetraspanin family. In several human cancers, the loss of KAI1 expression correlates with the increased metastatic potential of the tumour.

### MHC (major histocompatibility complex) class I molecules

Molecules that present peptides derived from the intracellular protein pool to T cells that eradicate cells displaying foreign peptides. Human cytomegalovirus (HCMV) uses the ERAD system to target MHC class I heavy chains (subunits of the MHC class I molecule) for destruction, to prevent the presentation of virus-derived peptides.

### Pael-R (Pael receptor)

A substrate of the ubiquitin ligase Parkin. When Parkin is inactivated, unfolded Pael-R accumulates in the ER and causes neuronal death, suggesting that Pael-R has an important role in the pathogenesis of autosomal recessive juvenile Parkinson's disease.

### Pdr5

An ATP-binding cassette transporter that confers multidrug resistance to the yeast *Saccharomyces cerevisiae*.

### Ribophorin I

A membrane-anchored subunit of the oligosaccharyltransferase complex. Deletion of the C-terminal transmembrane domain yields the misfolded, soluble RI322 variant, which bears a single N-linked glycan.

### Transthyretin (TTR)

A non-glycosylated serum protein secreted by the liver that functions as a carrier of the hormone thyroxine. Mutations that destabilize the homotetramer result in the formation of amyloid fibrils, causing neurodegeneration and organ failure.



## Adjusting the capacity of folding and disposal

Environmental changes and developmental processes can drastically change the workload of the ER. To maintain ER homeostasis, a coordinated programme known as the unfolded protein response (UPR) adjusts the folding capacity of the ER to meet demand and engages the destructive pathway if necessary<sup>94</sup>. The sensor proteins Ire1 (inositol requiring enzyme 1) in yeast and IRE1, PERK (dsRNA-activated protein kinase-like ER kinase) and ATF6 (activating transcription factor 6) in mammals detect elevated levels of unfolded polypeptides in the ER lumen and signal for increased expression of folding promoters and destructive modules. Moreover, cells have evolved additional strategies to save resources during times of stress. When the UPR is induced, selected messenger RNAs are degraded by a mechanism that requires IRE1 (ref. 95). Targets of this pathway include proteins that increase the workload of the ER without contributing to its fidelity. Removal of the signal sequence or the introduction of frame shifts stabilize mRNAs that are otherwise degraded, suggesting that the mRNA is processed at the ER membrane, perhaps by IRE1 itself. An additional pathway denies the entry of certain secretory proteins into the ER<sup>96</sup>. During acute stress, cargo proteins such as PrP (prion protein) are rerouted to the cytosol for immediate degradation by the proteasome, although BiP continues to enter the ER. A chimaeric PrP equipped with the BiP signal sequence can enter the ER of stressed cells, suggesting that this process is largely regulated by the protein's signal sequence, although the underlying structural differences are not clear. The mechanisms that regulate which proteins are admitted to the ER are collectively referred to as 'pre-emptive quality control'.

## Future perspectives

More and more pieces of the ERAD pathway are coming to light, yet even some basic questions remain unresolved. How do Htm1 and members of the EDEM family recognize their targets? Does Htm1 use Pdi1 for this purpose, and does calnexin confer specificity to the actions of EDEM1? Are other members of the EDEM family also associated with chaperones? Which mechanisms govern the recognition of misfolded proteins that bear no glycans? Perhaps an ERAD component binds to non-native proteins, and the probability of their breakdown increases with the duration or frequency of the interaction? The HRD ligase binds substrates through Hrd3. Is this contact critical for substrate selection? How are substrates handled after Yos9 has recognized the correct glycan code, and how does the HRD ligase release proteins that are not suitable for degradation? Which factors make up the pore that shuttles aberrant proteins into the cytosol? What is the function of the ubiquitin-binding domains found in several of the cytosolic ERAD components? Perhaps these domains contribute to the directionality of the ERAD pathway. This could be accomplished if downstream components have a higher affinity for ubiquitin than factors that act earlier in the pathway. With these questions in mind, it would be useful to have an *in vitro* system that recapitulates the breakdown of a misfolded glycoprotein. Now that most of the ERAD factors have been identified, as well as the glycan detected by Yos9, it might be possible to establish such a system. ■

- Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
- Schubert, U. *et al.* Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. *Nature* **404**, 770–774 (2000).
- Hampton, R. Y. Proteolysis and sterol regulation. *Annu. Rev. Cell Dev. Biol.* **18**, 345–378 (2002).
- Gurkan, C., Stagg, S. M., Lapointe, P. & Balch, W. E. The COPII cage: unifying principles of vesicle coat assembly. *Nature Rev. Mol. Cell Biol.* **7**, 727–738 (2006).
- Wiseman, R. L., Powers, E. T., Buxbaum, J. N., Kelly, J. W. & Balch, W. E. An adaptable standard for protein export from the endoplasmic reticulum. *Cell* **131**, 809–821 (2007).
- Daniels, R., Kurowski, B., Johnson, A. E. & Hebert, D. N. N-linked glycans direct the cotranslational folding pathway of influenza hemagglutinin. *Mol. Cell* **11**, 79–90 (2003).
- Molinari, M. N-glycan structure dictates extension of protein folding or onset of disposal. *Nature Chem. Biol.* **3**, 313–320 (2007).
- Solda, T., Galli, C., Kaufman, R. J. & Molinari, M. Substrate-specific requirements for UGT1-dependent release from calnexin. *Mol. Cell* **27**, 238–249 (2007). This study shows that some substrates are released from the calnexin cycle after one binding event, whereas others undergo several binding rounds.
- Hebert, D. N., Foellmer, B. & Helenius, A. Glucose trimming and reglucosylation determine

- glycoprotein association with calnexin in the endoplasmic reticulum. *Cell* **81**, 425–433 (1995).
- Helenius, A. & Aebi, M. Roles of N-linked glycans in the endoplasmic reticulum. *Annu. Rev. Biochem.* **73**, 1019–1049 (2004).
  - Camirand, A., Heysen, A., Grondin, B. & Herscovics, A. Glycoprotein biosynthesis in *Saccharomyces cerevisiae*. Isolation and characterization of the gene encoding a specific processing  $\alpha$ -mannosidase. *J. Biol. Chem.* **266**, 15120–15127 (1991).
  - Jelinek-Kelly, S. & Herscovics, A. Glycoprotein biosynthesis in *Saccharomyces cerevisiae*. Purification of the  $\alpha$ -mannosidase which removes one specific mannose residue from Man<sub>5</sub>GlcNAc. *J. Biol. Chem.* **263**, 14757–14763 (1988).
  - Jakob, C. A., Burda, P., Roth, J. & Aebi, M. Degradation of misfolded endoplasmic reticulum glycoproteins in *Saccharomyces cerevisiae* is determined by a specific oligosaccharide structure. *J. Cell Biol.* **142**, 1223–1233 (1998).
  - Jakob, C. A. *et al.* Htm1p, a mannosidase-like protein, is involved in glycoprotein degradation in yeast. *EMBO Rep.* **2**, 423–430 (2001).
  - Hosokawa, N. *et al.* A novel ER  $\alpha$ -mannosidase-like protein accelerates ER-associated degradation. *EMBO Rep.* **2**, 415–422 (2001).
  - Clerc, S. *et al.* Htm1 protein generates the N-glycan signal for glycoprotein degradation in the endoplasmic reticulum. *J. Cell Biol.* **184**, 159–172 (2009). This study identified Htm1 as a factor involved in the generation of mannose, oligosaccharides on ERAD substrates.
  - Olivari, S. & Molinari, M. Glycoprotein folding and the role of EDEM1, EDEM2 and EDEM3 in degradation of folding-defective glycoproteins. *FEBS Lett.* **581**, 3658–3664 (2007).
  - Tremblay, L. O. & Herscovics, A. Cloning and expression of a specific human  $\alpha$ 1,2-mannosidase that trims Man<sub>5</sub>GlcNAc<sub>2</sub> to Man<sub>5</sub>GlcNAc<sub>2</sub> isomer B during N-glycan biosynthesis. *Glycobiology* **9**, 1073–1078 (1999).
  - Avezov, E., Frenkel, Z., Ehrlich, M., Herscovics, A. & Lederkremer, G. Z. Endoplasmic reticulum (ER) mannosidase I is compartmentalized and required for N-glycan trimming to Man<sub>5</sub>GlcNAc<sub>2</sub> in glycoprotein ER-associated degradation. *Mol. Biol. Cell* **19**, 216–225 (2008).
  - Hosokawa, N. *et al.* Enhancement of endoplasmic reticulum (ER) degradation of misfolded Null Hong Kong  $\alpha$ 1-antitrypsin by human ER mannosidase I. *J. Biol. Chem.* **278**, 26287–26294 (2003).
  - Wu, Y., Swulius, M. T., Moremen, K. W. & Sifers, R. N. Elucidation of the molecular logic by which misfolded  $\alpha$ 1-antitrypsin is preferentially selected for degradation. *Proc. Natl Acad. Sci. USA* **100**, 8229–8234 (2003).
  - Hirao, K. *et al.* EDEM3, a soluble EDEM homolog, enhances glycoprotein endoplasmic reticulum-associated degradation and mannose trimming. *J. Biol. Chem.* **281**, 9650–9658 (2006).
  - Olivari, S. *et al.* EDEM1 regulates ER-associated degradation by accelerating de-mannosylation of folding-defective polypeptides and by inhibiting their covalent aggregation. *Biochem. Biophys. Res. Commun.* **349**, 1278–1284 (2006).
  - Moremen, K. W. & Molinari, M. N-linked glycan recognition and processing: the molecular basis of endoplasmic reticulum quality control. *Curr. Opin. Struct. Biol.* **16**, 592–599 (2006).
  - Molinari, M., Calanca, V., Galli, C., Lucca, P. & Paganetti, P. Role of EDEM in the release of misfolded glycoproteins from the calnexin cycle. *Science* **299**, 1397–1400 (2003).
  - Oda, Y., Hosokawa, N., Wada, I. & Nagata, K. EDEM as an acceptor of terminally misfolded glycoproteins released from calnexin. *Science* **299**, 1394–1397 (2003).
  - Carvalho, P., Goder, V. & Rapoport, T. A. Distinct ubiquitin-ligase complexes define convergent pathways for the degradation of ER proteins. *Cell* **126**, 361–373 (2006).
  - Vashist, S. & Ng, D. T. Misfolded proteins are sorted by a sequential checkpoint mechanism of ER quality control. *J. Cell Biol.* **165**, 41–52 (2004).
  - Bays, N. W., Gardner, R. G., Seelig, L. P., Joazeiro, C. A. & Hampton, R. Y. Hrd1p/Der3p is a membrane-anchored ubiquitin ligase required for ER-associated degradation. *Nature Cell Biol.* **3**, 24–29 (2001).
  - Deak, P. M. & Wolf, D. H. Membrane topology and function of Der3/Hrd1p as a ubiquitin-protein ligase (E3) involved in endoplasmic reticulum degradation. *J. Biol. Chem.* **276**, 10663–10669 (2001).
  - References 29 and 30 characterize Hrd1 as an E3 ligase that ubiquitylates several ERAD substrates.
  - Knop, M., Finger, A., Braun, T., Hellmuth, K. & Wolf, D. H. Der1, a novel protein specifically required for endoplasmic reticulum degradation in yeast. *EMBO J.* **15**, 753–763 (1996).
  - Buschhorn, B. A., Kostova, Z., Medicherla, B. & Wolf, D. H. A genome-wide screen identifies Yos9p as essential for ER-associated degradation of glycoproteins. *FEBS Lett.* **577**, 422–426 (2004).
  - Denic, V., Quan, E. M. & Weissman, J. S. A luminal surveillance complex that selects misfolded glycoproteins for ER-associated degradation. *Cell* **126**, 349–359 (2006).
  - Gauss, R., Jarosch, E., Sommer, T. & Hirsch, C. A complex of Yos9p and the HRD ligase integrates endoplasmic reticulum quality control into the degradation machinery. *Nature Cell Biol.* **8**, 849–854 (2006). References 33 and 34 show that Yos9 and Hrd3 act as a surveillance platform that binds to aberrant proteins and decodes the glycosylation status of a protein.
  - Quan, E. M. *et al.* Defining the glycan destruction signal for endoplasmic reticulum-associated degradation. *Mol. Cell* **32**, 870–877 (2008). This study characterizes N-glycans containing terminal  $\alpha$ 1,6-mannose as the preferred ligand of Yos9.
  - Biederer, T., Volkwein, C. & Sommer, T. Role of Cue1p in ubiquitination and degradation at the ER surface. *Science* **278**, 1806–1809 (1997).
  - Swanson, R., Locher, M. & Hochstrasser, M. A conserved ubiquitin ligase of the nuclear envelope/endoplasmic reticulum that functions in both ER-associated and Mada2 repressor degradation. *Genes Dev.* **15**, 2660–2674 (2001).
  - Kreft, S. G., Wang, L. & Hochstrasser, M. Membrane topology of the yeast endoplasmic reticulum-localized ubiquitin ligase Doa10 and comparison with its human ortholog TEB4 (MARCVI). *J. Biol. Chem.* **281**, 4646–4653 (2006).
  - Fang, S. *et al.* The tumor autocrine motility factor receptor, gp78, is a ubiquitin protein ligase implicated in degradation from the endoplasmic reticulum. *Proc. Natl Acad. Sci. USA* **98**, 14422–14427 (2001).

40. Kikkert, M. *et al.* Human HRD1 is an E3 ubiquitin ligase involved in degradation of proteins from the endoplasmic reticulum. *J. Biol. Chem.* **279**, 3525–3534 (2004).
41. Nadav, E. *et al.* A novel mammalian endoplasmic reticulum ubiquitin ligase homologous to the yeast Hrd1. *Biochem. Biophys. Res. Commun.* **303**, 91–97 (2003).
42. Mueller, B., Lilley, B. N. & Ploegh, H. L. SEL1L, the homologue of yeast Hrd3p, is involved in protein dislocation from the mammalian ER. *J. Cell Biol.* **175**, 261–270 (2006).
43. Kokame, K., Agarwala, K. L., Kato, H. & Miyata, T. Herp, a new ubiquitin-like membrane protein induced by endoplasmic reticulum stress. *J. Biol. Chem.* **275**, 32846–32853 (2000).
44. Schulze, A. *et al.* The ubiquitin-domain protein HERP forms a complex with components of the endoplasmic reticulum associated degradation pathway. *J. Mol. Biol.* **354**, 1021–1027 (2005).
45. Lilley, B. N. & Ploegh, H. L. A membrane protein required for dislocation of misfolded proteins from the ER. *Nature* **429**, 834–840 (2004).
46. Ye, Y., Shibata, Y., Yun, C., Ron, D. & Rapoport, T. A. A membrane protein complex mediates retro-translocation from the ER lumen into the cytosol. *Nature* **429**, 841–847 (2004).  
**References 45 and 46 show that Derlin-1 is required for the membrane extraction of certain aberrant proteins from the ER.**
47. Bernasconi, R., Pertel, T., Luban, J. & Molinari, M. A dual task for the Xbp1-responsive OS-9 variants in the mammalian endoplasmic reticulum: inhibiting secretion of misfolded protein conformers and enhancing their disposal. *J. Biol. Chem.* **283**, 16446–16454 (2008).
48. Christianson, J. C., Shaler, T. A., Tyler, R. E. & Kopito, R. R. OS-9 and GRP94 deliver mutant  $\alpha$ 1-antitrypsin to the Hrd1–SEL1L ubiquitin ligase complex for ERAD. *Nature Cell Biol.* **10**, 272–282 (2008).
49. Hosokawa, N. *et al.* Human XTP3-B forms an ER quality control scaffold with the HRD1–SEL1L ubiquitin ligase complex and BiP. *J. Biol. Chem.* **283**, 20914–20924 (2008).
50. Kaneko, M., Ishiguro, M., Niinuma, Y., Uesugi, M. & Nomura, Y. Human HRD1 protects against ER stress-induced apoptosis through ER-associated degradation. *FEBS Lett.* **532**, 147–152 (2002).
51. Cattaneo, M. *et al.* SEL1L and HRD1 are involved in the degradation of unassembled secretory Ig- $\mu$  chains. *J. Cell. Physiol.* **215**, 794–802 (2008).
52. Okuda-Shimizu, Y. & Hendershot, L. M. Characterization of an ERAD pathway for nonglycosylated BiP substrates, which require Herp. *Mol. Cell* **28**, 544–554 (2007).
53. Chen, B. *et al.* The activity of a human endoplasmic reticulum-associated degradation E3, gp78, requires its Cue domain, RING finger, and an E2-binding site. *Proc. Natl Acad. Sci. USA* **103**, 341–346 (2006).
54. Shen, Y., Ballar, P. & Fang, S. Ubiquitin ligase gp78 increases solubility and facilitates degradation of the Z variant of  $\alpha$ 1-antitrypsin. *Biochem. Biophys. Res. Commun.* **349**, 1285–1293 (2006).
55. Song, B. L., Sever, N. & DeBose-Boyd, R. A. Gp78, a membrane-anchored ubiquitin ligase, associates with Insig-1 and couples sterol-regulated ubiquitination to degradation of HMG CoA reductase. *Mol. Cell* **19**, 829–840 (2005).
56. Tsai, Y. C. *et al.* The ubiquitin ligase gp78 promotes sarcoma metastasis by targeting KAI1 for degradation. *Nature Med.* **13**, 1504–1509 (2007).
57. Hassink, G. *et al.* TEB4 is a C4HC3 RING finger-containing ubiquitin ligase of the endoplasmic reticulum. *Biochem. J.* **388**, 647–655 (2005).
58. Younger, J. M. *et al.* Sequential quality-control checkpoints triage misfolded cystic fibrosis transmembrane conductance regulator. *Cell* **126**, 571–582 (2006).
59. McDonough, H. & Patterson, C. CHIP: a link between the chaperone and proteasome systems. *Cell Stress Chaperones* **8**, 303–308 (2003).
60. Wang, B. *et al.* BAP31 interacts with Sec61 translocons and promotes retrotranslocation of CFTR $\Delta$ 508 via the Derlin-1 complex. *Cell* **133**, 1080–1092 (2008).
61. Kitada, T. *et al.* Mutations in the Parkin gene cause autosomal recessive juvenile parkinsonism. *Nature* **392**, 605–608 (1998).
62. Imai, Y. *et al.* CHIP is associated with Parkin, a gene responsible for familial Parkinson's disease, and enhances its ubiquitin ligase activity. *Mol. Cell* **10**, 55–67 (2002).
63. Yoshida, Y. *et al.* E3 ubiquitin ligase that recognizes sugar chains. *Nature* **418**, 438–442 (2002).
64. Plemper, R. K., Bohmler, S., Bordallo, J., Sommer, T. & Wolf, D. H. Mutant analysis links the translocon and BiP to retrograde protein transport for ER degradation. *Nature* **388**, 891–895 (1997).
65. Wiertz, E. J. *et al.* Sec61-mediated transfer of a membrane protein from the endoplasmic reticulum to the proteasome for destruction. *Nature* **384**, 432–438 (1996).
66. Ploegh, H. L. A lipid-based model for the creation of an escape hatch from the endoplasmic reticulum. *Nature* **448**, 435–438 (2007).
67. Gauss, R., Sommer, T. & Jarosch, E. The Hrd1p ligase complex forms a linchpin between ER-lumenal substrate selection and Cdc48p recruitment. *EMBO J.* **25**, 1827–1835 (2006).
68. Hirsch, C., Blom, D. & Ploegh, H. L. A role for N-glycanase in the cytosolic turnover of glycoproteins. *EMBO J.* **22**, 1036–1046 (2003).
69. Suzuki, T., Park, H., Hollingsworth, N. M., Sternglanz, R. & Lennarz, W. J. PNG1, a yeast gene encoding a highly conserved peptide:N-glycanase. *J. Cell Biol.* **149**, 1039–1052 (2000).
70. Lilley, B. N. & Ploegh, H. L. Multiprotein complexes that link dislocation, ubiquitination, and extraction of misfolded proteins from the endoplasmic reticulum membrane. *Proc. Natl Acad. Sci. USA* **102**, 14296–14301 (2005).
71. Oda, Y. *et al.* Derlin-2 and Derlin-3 are regulated by the mammalian unfolded protein response and are required for ER-associated degradation. *J. Cell Biol.* **172**, 383–393 (2006).
72. de Virgilio, M., Weninger, H. & Ivessa, N. E. Ubiquitination is required for the retro-translocation of a short-lived luminal endoplasmic reticulum glycoprotein to the cytosol for degradation by the proteasome. *J. Biol. Chem.* **273**, 9734–9743 (1998).
73. Shamu, C. E., Flierman, D., Ploegh, H. L., Rapoport, T. A. & Chau, V. Polyubiquitination is required for US11-dependent movement of MHC class I heavy chain from endoplasmic reticulum into cytosol. *Mol. Biol. Cell* **12**, 2546–2555 (2001).
74. Bays, N. W., Wilhovsky, S. K., Goradia, A., Hodgkiss-Harlow, K. & Hampton, R. Y. HRD4/NPL4 is required for the proteasomal processing of ubiquitinated ER proteins. *Mol. Biol. Cell* **12**, 4114–4128 (2001).
75. Braun, S., Matuschewski, K., Rape, M., Thoms, S. & Jentsch, S. Role of the ubiquitin-selective CDC48(UFD1/NPL4) chaperone (segregase) in ERAD of OLE1 and other substrates. *EMBO J.* **21**, 615–621 (2002).
76. Jarosch, E. *et al.* Protein dislocation from the ER requires polyubiquitination and the AAA-ATPase Cdc48. *Nature Cell Biol.* **4**, 134–139 (2002).
77. Ye, Y., Meyer, H. H. & Rapoport, T. A. The AAA ATPase Cdc48/p97 and its partners transport proteins from the ER into the cytosol. *Nature* **414**, 652–656 (2001).  
**References 74–77 demonstrate that yeast Cdc48, or p97 in mammals, mobilizes ERAD substrates from the ER membrane.**
78. Zhang, X. *et al.* Structure of the AAA ATPase p97. *Mol. Cell* **6**, 1473–1484 (2000).
79. DeLaBarre, B., Christianson, J. C., Kopito, R. R. & Brunger, A. T. Central pore residues mediate the p97/VCP activity required for ERAD. *Mol. Cell* **22**, 451–462 (2006).
80. Neuber, O., Jarosch, E., Volkwein, C., Walter, J. & Sommer, T. Ubx2 links the Cdc48 complex to ER-associated protein degradation. *Nature Cell Biol.* **7**, 993–998 (2005).
81. Schubert, C. & Buchberger, A. Membrane-bound Ubx2 recruits Cdc48 to ubiquitin ligases and their substrates to ensure efficient ER-associated protein degradation. *Nature Cell Biol.* **7**, 999–1006 (2005).
82. Liang, J. *et al.* Characterization of erasin (UBXD2): a new ER protein that promotes ER-associated protein degradation. *J. Cell Sci.* **119**, 4011–4024 (2006).
83. Medicherla, B., Kostova, Z., Schaefer, A. & Wolf, D. H. A genomic screen identifies Dsk2p and Rad23p as essential components of ER-associated degradation. *EMBO Rep.* **5**, 692–697 (2004).
84. Richly, H. *et al.* A series of ubiquitin binding factors connects CDC48/p97 to substrate multiubiquitylation and proteasomal targeting. *Cell* **120**, 73–84 (2005).  
**This study demonstrates how components of the ubiquitin-proteasome system cooperate to deliver ubiquitylated proteins to the proteasome.**
85. Deveraux, Q., Jensen, C. & Rechsteiner, M. Molecular cloning and expression of a 26S protease subunit enriched in dileucine repeats. *J. Biol. Chem.* **270**, 23726–23729 (1995).
86. Husnjak, K. *et al.* Proteasome subunit Rpn13 is a novel ubiquitin receptor. *Nature* **453**, 481–488 (2008).
87. Schreiner, P. *et al.* Ubiquitin docking at the proteasome through a novel pleckstrin-homology domain interaction. *Nature* **453**, 548–552 (2008).
88. Suzuki, T., Park, H., Kwofie, M. A. & Lennarz, W. J. Rad23 provides a link between the Png1 deglycosylating enzyme and the 26S proteasome in yeast. *J. Biol. Chem.* **276**, 21601–21607 (2001).
89. Blom, D., Hirsch, C., Stern, P., Tortorella, D. & Ploegh, H. L. A glycosylated type I membrane protein becomes cytosolic when peptide:N-glycanase is compromised. *EMBO J.* **23**, 650–658 (2004).
90. Misaghi, S., Pacold, M., Blom, D., Ploegh, H. L. & Korb, G. A. Using a small molecule inhibitor of peptide:N-glycanase to probe its role in glycoprotein turnover. *Chem. Biol.* **11**, 1677–1687 (2004).
91. Hu, M. *et al.* Structure and mechanisms of the proteasome-associated deubiquitinating enzyme USP14. *EMBO J.* **24**, 3747–3756 (2005).
92. Verma, R. *et al.* Role of Rpn11 metalloprotease in deubiquitination and degradation by the 26S proteasome. *Science* **298**, 611–615 (2002).
93. Yao, T. & Cohen, R. E. A cryptic protease couples deubiquitination and degradation by the proteasome. *Nature* **419**, 403–407 (2002).
94. Ron, D. & Walter, P. Signal integration in the endoplasmic reticulum unfolded protein response. *Nature Rev. Mol. Cell Biol.* **8**, 519–529 (2007).
95. Hollien, J. & Weissman, J. S. Decay of endoplasmic reticulum-localized mRNAs during the unfolded protein response. *Science* **313**, 104–107 (2006).
96. Kang, S. W. *et al.* Substrate-specific translocational attenuation during ER stress defines a pre-emptive quality control pathway. *Cell* **127**, 999–1013 (2006).
97. Hammarstrom, P. *et al.* D18G transthyretin is monomeric, aggregation prone, and not detectable in plasma and cerebrospinal fluid: a prescription for central nervous system amyloidosis? *Biochemistry* **42**, 6656–6663 (2003).
98. Sekijima, Y. *et al.* Energetic characteristics of the new transthyretin variant A25T may explain its atypical central nervous system pathology. *Lab. Invest.* **83**, 409–417 (2003).
99. Sekijima, Y. *et al.* The biological and chemical basis for tissue-selective amyloid disease. *Cell* **121**, 73–85 (2005).

**Acknowledgements** We thank all members of the Sommer laboratory for critical reading of the manuscript. We apologize to those authors whose work could not be cited because of space limitations. Work in T.S.'s laboratory is supported by the German Research Foundation (DFG) and the RUBICON Network of Excellence. R.G. was a fellow of the Boehringer Ingelheim Fonds. S.C.H. is the recipient of a Helmholtz PhD fellowship.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Correspondence should be addressed to T.S. ([tsommer@mdc-berlin.de](mailto:tsommer@mdc-berlin.de)).



# Principles of ubiquitin and SUMO modifications in DNA repair

Steven Bergink<sup>1</sup> & Stefan Jentsch<sup>1</sup>

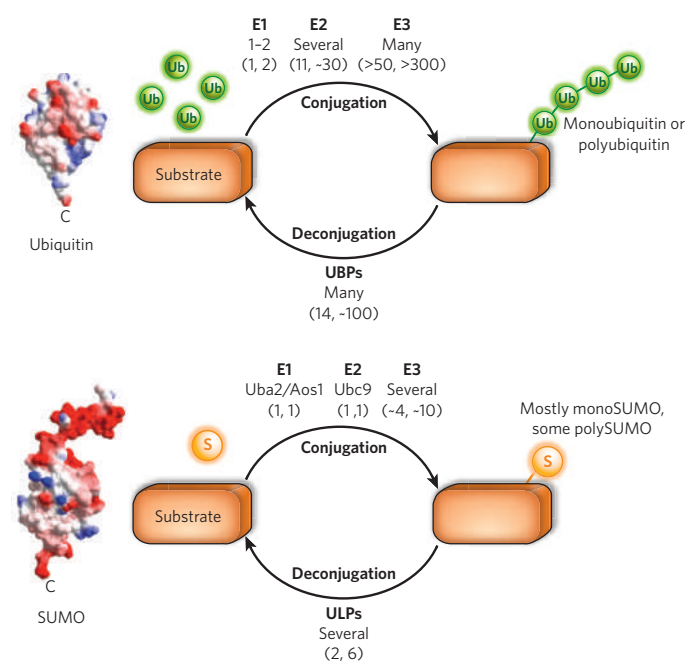
**With the discovery in the late 1980s that the DNA-repair gene *RAD6* encodes a ubiquitin-conjugating enzyme, it became clear that protein modification by ubiquitin conjugation has a much broader significance than had previously been assumed. Now, two decades later, ubiquitin and its cousin SUMO are implicated in a range of human diseases, including breast cancer and Fanconi anaemia, giving fresh momentum to studies focused on the relationships between ubiquitin, SUMO and DNA-repair pathways.**

Modification of proteins by conjugation of ubiquitin (ubiquitylation), or its relative SUMO (SUMOylation), is central to most aspects of eukaryotic life. As for phosphorylation, these modifications are reversible, making them ideally suited for regulatory purposes. The conjugation of these modifiers is ATP dependent and involves activating enzymes (E1), conjugating enzymes (E2) and ligases (E3)<sup>1</sup>. In the ubiquitin pathway, substrate specificity is usually provided by the E3 ligases, which typically have substrate-binding sites, and sometimes by a combination of E3 ligases with E2 enzymes<sup>2</sup>. To achieve high specificity, cells express many different E2s and even more E3s, and these enzymes are often targets for regulation (Fig. 1). Surprisingly, although SUMOylation also affects a large number of substrates, its pathway is much simpler and involves an E1, a single E2 (Ubc9) and only a few E3 ligases (Fig. 1). Ubc9 usually binds the substrate directly, but the SUMO E3s also seem to contribute to substrate specificity. SUMOylation often targets a lysine residue within a 'consensus' sequence, but other lysine residues can be modified as well. In contrast to yeast, vertebrates have three SUMO variants.

Complexity arises because both ubiquitin and SUMO can form poly-modifier 'chains' — that is, they can become conjugated to lysine residues of another (already substrate-conjugated) modifier molecule. In contrast to polyubiquitin chains, little is known about the general function of polySUMO chains (yeast SUMO and mammalian SUMO-2/3 can form chains but SUMO-1 cannot). In the case of the protein PML, polySUMOylation recruits the ubiquitin ligase RNF4 for polyubiquitylation and protein degradation<sup>3</sup>. Further complication came to light with the discovery that branched, 'tree-like' polyubiquitin modifications occur *in vivo* (when multiple ubiquitin lysines are used)<sup>4</sup> and that SUMO may be modified by ubiquitin, possibly resulting in mixed ubiquitin–SUMO chains<sup>3</sup>, although the relevance of these modifications is currently unclear. Importantly, both ubiquitylation and SUMOylation are reversible, as hydrolases can remove the modifications completely or trim the modifier chains or trees to a certain size.

The available structural data suggest that the modifiers may introduce conformational changes of the substrates, but they usually barely affect the folding of the conjugated substrate around the attachment site. This is probably due to the fact that the carboxy-terminal glycine residues of the modifiers are positioned on flexible protruding tails. However, because the modifiers are proteins, they often provide surfaces for protein interactions, or, because of their bulkiness, prevent or loosen protein–protein interactions. Downstream factors, partners that bind to the various types of modifications, or factors that

are repelled by these modifications are therefore crucial. Indeed, many different sequence motifs have been identified that represent binding modules specific for ubiquitin<sup>5</sup> (such as ubiquitin-associated (UBA) domains, ubiquitin-interacting motifs (UIM) or ubiquitin-binding zinc finger (UBZ) domains) or SUMO-interacting motifs (SIM)<sup>6</sup>. Remarkably, some of them can even distinguish between the different types of poly-modifier linkages<sup>7</sup>.



**Figure 1 | Schematic diagram of ubiquitin and SUMO modification pathways.** Ubiquitin (Ub, top pathway) and SUMO (S, bottom), known as modifiers, are small polypeptides that are usually conjugated to internal lysine residues of target proteins through their C-terminal glycine residues, forming isopeptide bonds. Both modifications are reversible. Modification can be by a single modifier or by a chain of covalently linked modifiers (polyubiquitylation or polySUMOylation). The number of enzymes in the yeast *Saccharomyces cerevisiae* and in humans, respectively, are shown in parentheses. UBPs, ubiquitin-specific proteases; ULPs, ubiquitin-like protein (SUMO)-specific proteases.

<sup>1</sup>Department of Molecular Cell Biology, Max Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany.

## Regulation of repair by ubiquitylation

Although early studies suggested that the central function of ubiquitin might be proteasomal degradation, it is now becoming clear that protein modification by ubiquitylation or SUMOylation has a much broader significance. Especially eye-catching is the fact that various nuclear functions seem to be controlled by ubiquitylation and SUMOylation. Moreover, mass spectroscopy has identified prominent nuclear proteins as substrates for these modifiers, most notably for SUMO<sup>8</sup>. Strikingly, all the major DNA-repair pathways, damage-avoidance mechanisms and checkpoint responses are regulated in some way by ubiquitylation, SUMOylation or both (Table 1). Indeed, genes of the RAD6-damage tolerance pathway encode mostly enzymes involved in the ubiquitin pathway. Similarly, the interstrand crosslink-repair pathway linked to Fanconi anaemia seems to consist largely of enzymes and substrates of the ubiquitin system.

DNA can be damaged in different ways and lesions can arise through exogenous and endogenous damage. The type of damage and the cell-cycle stage at which it occurs determine the pathway used for repair<sup>9</sup> (Box 1). DNA damage usually induces the activation of DNA repair or damage-avoidance pathways, and often a checkpoint response that triggers cell-cycle arrest to make time for the repair. This reaction, referred to as the DNA-damage response (DDR)<sup>10,11</sup>, is typically initiated by proteins that recognize DNA lesions, and is followed by the recruitment and activation of proteins that trigger checkpoint signalling or directly perform the necessary repair steps. In some cases, repair takes place within large protein assemblies (for recombinational repair, for example), which are microscopically identifiable as repair foci. When the repair is completed, the machinery needs to be disassembled and the DDR turned off.

DDR pathways provide startling cases of regulation by ubiquitin

and SUMO, as every facet of these modifiers seems to be involved. Moreover, some of these pathways provide illuminating examples of how these two modification systems can talk to each other and coordinate these complex events. Here we focus on a few illuminating cases and reflect on the basic mechanisms and principles of modification by ubiquitin and SUMO. Given their relationship to diseases such as breast cancer, Fanconi anaemia and xeroderma pigmentosum, the ways ubiquitin and SUMO affect DNA-repair pathways has become something of a hot topic.

## Regulation of nucleotide-excision repair

Most DDR pathways are controlled by the ubiquitin–proteasome pathway (UPS). Examples are the degradation of stalled RNA polymerases at damaged DNA templates<sup>12</sup>, or the degradation of repair proteins after completed repair (see below). A case that merits special attention is found within the nucleotide-excision repair (NER) pathway (see Box 1). NER, perhaps best known for its ability to repair ultraviolet-light-induced lesions, also acts on most helix-distorting lesions. Lesions of this type are recognized by the XPC protein (xeroderma pigmentosum group C) complex<sup>13</sup> and the ultraviolet light, DNA-damage-binding (UV-DDB) complex<sup>14</sup>. When the lesion is recognized, the DNA surrounding the site is melted by helicases of the TFIIH complex. When the lesion-containing oligonucleotide has been excised, DNA polymerases fill in the resulting gap. XPC, encoded by one of the genes mutated in patients with the DNA-repair disorder xeroderma pigmentosum, forms a complex with the calcium-binding EF-hand centrin-2 protein (Cdc31 in yeast) and with either of the two related human proteins, RAD23A and RAD23B<sup>15,16</sup>. These two relatives of the yeast escort factor Rad23 have C-terminal ubiquitin-binding modules (UBA domains), which recognize ubiquitin–protein conjugates, and an N-terminal

**Table 1 | Examples of DNA-repair proteins modified by ubiquitin and SUMO**

Protein	Modification	Repair pathway	Function	Organism	References
PCNA	Ubiquitin	RRR RAD6 pathway, PRR	Translesion synthesis, error-free pathway	<i>Saccharomyces cerevisiae</i> , <i>Schizosaccharomyces pombe</i> , <i>Xenopus laevis</i> , birds, mammals	28, 62–64
PCNA	SUMO	RRR	Inhibition of recombination, inhibition of cohesion by preventing Eco1 from PCNA binding	<i>S. cerevisiae</i> , birds	28, 62–64
9-1-1	Ubiquitin	DDR	DNA-damage-induced transcription, 'SOS response'	<i>S. cerevisiae</i>	65
XPC (Rad4)	Ubiquitin	NER	Affinity change for DNA	Mammals, <i>S. cerevisiae</i>	21, 66
XPC	SUMO	NER	Unknown	Mammals	66
DDB2	Ubiquitin	NER	Proteasomal degradation	Mammals	23
TDG	SUMO	BER	Dissociation from AP site	Mammals	39
Rad52	SUMO	HR	Extranucleolar foci formation, protection against degradation	<i>S. cerevisiae</i> , mammals	42, 43
Rad54	Ubiquitin	HR	Degradation in G1	<i>S. pombe</i>	67
H2A	Ubiquitin	HR, NER	Unknown, recruitment of RAP80?	Mammals	45, 47, 48
XRCC4	SUMO	NHEJ	Change in localization	Mammals	68
Ku70	SUMO	NHEJ	Unknown	<i>S. cerevisiae</i>	69
Ku80	Ubiquitin	NHEJ	Removal from DNA	<i>X. laevis</i>	70
BLM	SUMO	RRR	Change in localization	Mammals	71
WRN	SUMO	RRR	Unknown	Mammals	72
Sgs1	SUMO	RRR	Unknown	<i>S. cerevisiae</i>	73
TopI	SUMO	RRR	Unknown	<i>S. cerevisiae</i>	74
TopII	SUMO	RRR	Regulation of centromeric cohesion	<i>S. cerevisiae</i> , mammals	74, 75
FANCD2	Ubiquitin	FA	Targeting to chromatin, protein interactions	Mammals	55
FANCI	Ubiquitin	FA	Targeting to chromatin, protein interactions	Mammals	57
RNA Pol II	Ubiquitin	Transcription	Degradation upon polymerase stalling	<i>S. cerevisiae</i>	12

BER, base excision repair; DDR, DNA-damage response; FA, Fanconi's-anaemia pathway; HR, homologous recombination; NER, nucleotide-excision repair; NHEJ, non-homologous end-joining; PRR, postreplication repair; RRR, replication/recombination/repair.



ubiquitin-like domain, which normally targets the protein together with the bound ubiquitin conjugate to the proteasome<sup>17</sup>.

How ubiquitylation and the proteasome are involved in NER is a matter of debate, but several findings suggest non-canonical roles. In yeast, the XPC-related Rad4 protein is ubiquitylated and degraded by proteasomes<sup>18</sup>. Unexpectedly, however, in contrast to other known proteasomal pathways, Rad23 seems to protect Rad4 (and XPC) from degradation<sup>19,20</sup>. In humans, XPC ubiquitylation is carried out by a specific, cullin-based ubiquitin-ligase complex (CUL4a–ROC1–DDB1; UV–DDB)<sup>21</sup>. This ligase is normally kept in check by association with the COP9 signalosome (a particle related to the 'lid complex' of the 19S cap of the proteasome), which represses its ligase activity<sup>22</sup>. However, the signalosome dissociates from the ligase at DNA lesions, whereupon the two lesion-recognition proteins XPC and DDB2 are ubiquitylated. The consequences of the modification for the two substrates are different, however, as DDB2 polyubiquitylation triggers its destruction<sup>23</sup>, whereas XPC polyubiquitylation does not<sup>21</sup> (Fig. 2). Polyubiquitylated XPC has a higher affinity for DNA (both damaged and undamaged) *in vitro*, suggesting a regulatory role<sup>21</sup>. Whether these very different fates are caused by different types of ubiquitin chain (for example, linked at the lysine at position 48 (Lys 48) versus Lys 63) is not known, but an attractive idea is that polyubiquitylated XPC, perhaps in complex with RAD23 proteins, fosters the formation of specific protein assemblies needed for efficient repair.

Studies in yeast suggest a non-proteolytic role for the 19S cap of the proteasome in NER<sup>24</sup>. This is surprising as the 19S cap, usually part of the complete 26S proteasome, functions normally in substrate unfolding before substrate degradation. Because the 19S cap contains AAA-type ATPases<sup>25</sup> it has been suggested that the 19S particle might function in NER by remodelling repair-protein complexes<sup>24</sup>. Although the precise UPS-driven events in NER are far from clear, it is reasonable to speculate that perhaps a non-proteolytic function (for example, remodelling) is followed by proteasomal degradation, conceivably to clear the repair site of superseded repair proteins. As in other pathways<sup>26</sup>, these two steps may be triggered by two types of ubiquitin modification, for example, mono- or oligoubiquitylation followed by (Lys-48-linked) polyubiquitylation.

### A ubiquitin–SUMO switchboard

One of the most striking cases of control by ubiquitin and SUMO is proliferating cell nuclear antigen (PCNA). This homotrimeric, ring-shaped protein that encircles DNA acts as a processing factor for DNA polymerases and as a moving platform for factors that mediate replication-linked functions (such as chromatin assembly or sister-chromatid cohesion)<sup>27</sup>. Remarkably, PCNA can be modified on the same conserved lysine residue (Lys 164) either by monoubiquitylation, Lys-63-linked polyubiquitin chains, or SUMO (ref. 28) (Fig. 3). Both ubiquitylation and SUMOylation of PCNA occur in S phase, but ubiquitylation specifically occurs when DNA is damaged. Unexpectedly, about half of the known genes of the yeast *RAD6* DNA-damage-tolerance pathway encode enzymes that catalyse PCNA ubiquitylation. Among these, Rad18 and Rad5 are E3 ubiquitin ligases of the RING-finger class<sup>29</sup>. Rad6 itself is a conventional E2 ubiquitin-conjugating enzyme<sup>30</sup>, whereas Ubc13 and Mms2 (the latter lacks the active-site cysteine residue) form a heterodimeric E2 with the unique property of catalysing attachment of Lys-63-linked polyubiquitin chains. Rad6 and Rad18 catalyse PCNA monoubiquitylation, whereas Ubc13/Mms2 and Rad5 are needed to extend the modification by a Lys-63-linked polyubiquitin chain. The two ligases, Rad18 and Rad5, can heterodimerize, and both bind PCNA directly<sup>28,29</sup>.

During S phase, damaged DNA can lead to stalled replication forks, which in turn generate single-stranded DNA segments. These are subsequently coated by the heterotrimeric RPA protein. RPA then recruits Rad18 to these sites<sup>31</sup> and, owing to the affinity of Rad18 for PCNA and Rad6, monoubiquitylates PCNA on its Lys 164 residue<sup>28</sup>. PCNA monoubiquitylation mediates translesion synthesis (TLS), that is, DNA synthesis across the damaged site. TLS is catalysed by a set of specialized polymerases that can accommodate various bulky lesions

### Box 1 | DNA repair mechanisms

#### Non-homologous end joining (NHEJ)

As the name implies, this pathway repairs double-strand breaks (DSBs) without using the information of the homologous sequences. This pathway is error prone. After the broken ends are processed, the two ends are relinked.

#### Homologous recombination (HR)

Here the information on the sister chromatid is usually used as a template for the repair of a broken chromatid. After resection of the DSB, a 3' overhang invades the base pairs within the homologous sequence. DNA synthesis restores the missing information, after which the complex is disassembled, resulting in a repaired chromatid.

#### Interstrand crosslink repair

In metazoans, this is orchestrated by the Fanconi's anaemia pathway. Several repair and damage-avoidance pathways (HR, NER and PRR) are integrated to resolve the interstrand crosslink.

#### Base-excision repair (BER)

This deals with lesions that involve only one strand and mildly affect base pairing. A group of glycosidases with overlapping substrate affinity initiates the reaction by removing the affected base. A nuclease mediates strand incision, leaving an abasic site. The subsequent gap is filled in by polymerase- $\beta$ .

#### Nucleotide-excision repair (NER)

Lesions that cause more serious helix distortion but only affect one strand are dealt with by this pathway. After lesion recognition, the helix is unwound, followed by excision of the lesion-containing oligonucleotide. Finally, gap synthesis and ligation result in successful repair.



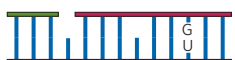
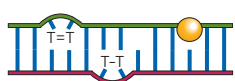

#### Mismatch repair (MMR)

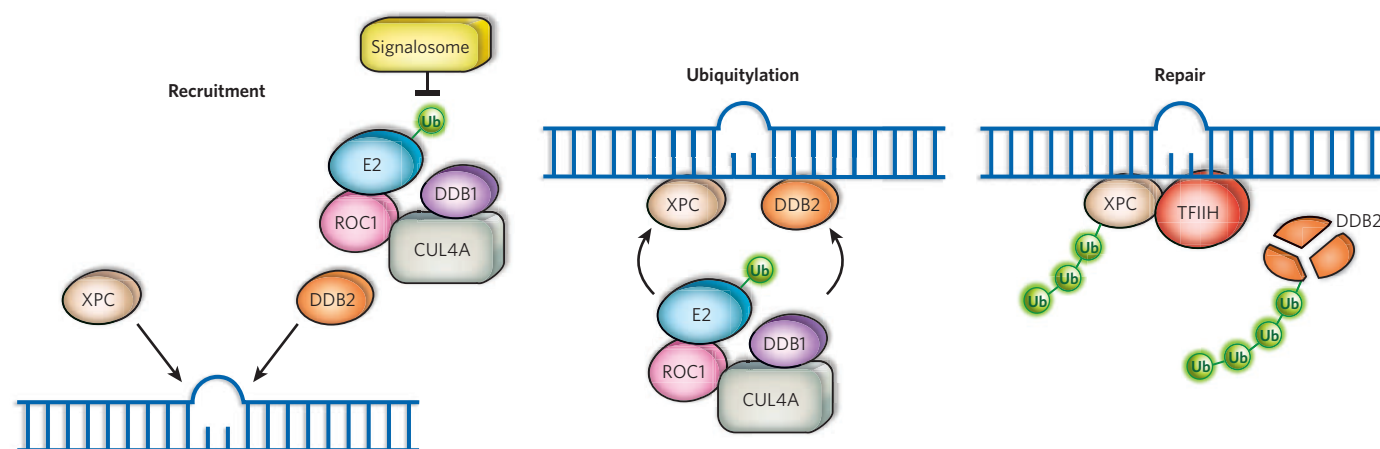
This restores mistakes in replication (wrong incorporations of nucleotides or even small gaps). The mismatch is recognized and removed by the action of several different proteins. The replication machinery then makes a second attempt, using the original strand as a template.

#### Postreplication repair (PRR)

Lesions that are encountered by the DNA polymerase are resolved by either translesion synthesis (TLS) or the error-free template switch combined with recombination. The modifications on the proliferating cell nuclear antigen (PCNA) ring determine the choice between these subpathways. Monoubiquitylation of a PCNA results in the use of low-fidelity polymerases that can bypass the lesion. Polyubiquitylation of PCNA evokes an error-free pathway that is less well understood.

#### Pathways involved in avoiding or repairing damage by lesion type

	Type of lesion	Pathway
	Double-strand breaks	NHEJ, HR
	Interstrand crosslinks	HR, PRR, NER
	Single-strand breaks, abasic sites, uracil	BER
	Cyclobutane pyrimidine dimer, (6-4) photoproduct, bulky adduct	PRR, NER
	AG mismatch, TC mismatch	MMR



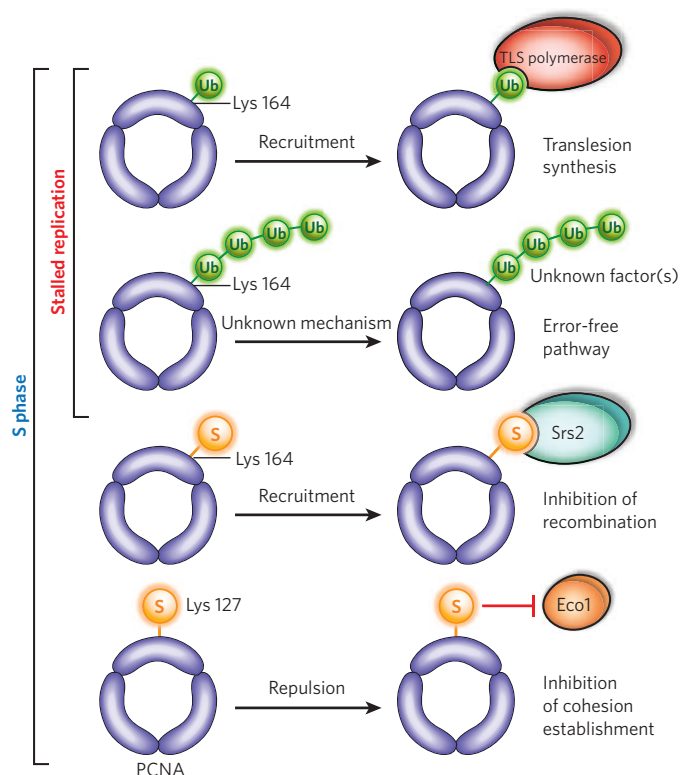
**Figure 2 | Scheme for XPC and DDB2 ubiquitylation.** Helix-distorting lesions are substrates for XPC and DDB2 (left). XPC binds to the undamaged strands of helix-distorted lesions (middle). DDB2 becomes part of a larger, cullin 4a-based E3 ubiquitin ligase complex (CUL4a, DDB1, ROC1 and an E2 ubiquitin-conjugating enzyme). In the absence of DNA damage, the ubiquitylation activity is repressed by the

signalosome (COP9) complex (upper left). Upon DNA damage, the signalosome dissociates, allowing the DDB2–E3 ligase complex to bind to the damaged site. Both XPC and DDB2 are substrates of the ubiquitin ligase; however, whereas DDB2 is degraded, XPC is not (right). The role of XPC ubiquitylation is currently not known, but it may promote specific protein interactions.

at their active sites<sup>32</sup>. Because of this feature, TLS polymerases are error-prone and can also insert an incorrect nucleotide, making TLS mutagenic. Remarkably, several TLS polymerases have ubiquitin-binding motifs (for example, UBM and UBZ domains), which may enable them to interact more specifically with monoubiquitylated PCNA<sup>33,34</sup>. Polyubiquitylation may happen if TLS fails (perhaps depending on the lesion) and results in a recombination-related error-free bypass. Precisely how this pathway works and whether polyubiquitylated PCNA recruits a specific factor are currently unknown.

Yeast PCNA can also be SUMOylated on Lys 164, and to a minor extent on Lys 127 (ref. 28). Both lysine residues are located on the outside rim of the ring-shaped molecule, away from the encircled DNA. SUMOylated PCNA specifically recruits Srs2, a helicase-like enzyme,

which potentially strips the recombinase Rad51 from chromatin<sup>35,36</sup>. This mechanism may aid replication by preventing unwanted recombination between the newly formed sister chromatids. Notably, Srs2 can bind PCNA directly, but, thanks to a C-terminal SUMO-interacting domain (SIM), it binds SUMOylated PCNA with much higher affinity<sup>35,36</sup>. The second SUMOylation site, Lys 127, lies directly at the binding pocket for most PCNA-interacting proteins (PIP-box proteins)<sup>28</sup>. Thus Lys 127 SUMOylation (and to a lesser extent Lys 164 SUMOylation) seems to prevent interaction with proteins that bind to this pocket through their PIP boxes<sup>27</sup>. One such factor is Eco1, which binds PCNA via a PIP-box and establishes sister-chromatid cohesion in S phase<sup>37</sup>. Unexpectedly, Eco1 binding to PCNA is prevented not only by PCNA SUMOylation but also by the binding of Ubc9 to PCNA<sup>37</sup>. These findings suggest a



**Figure 3 | The PCNA ubiquitin-SUMO switchboard.** PCNA (purple ring) can be modified at Lys 164 by (from the top) monoubiquitin, a Lys-63-linked polyubiquitin chain, or by SUMO. PCNA of *S. cerevisiae* can also be modified by SUMO at Lys 127. All modifications occur in S phase. Ubiquitylation occurs only where there is DNA damage or stalled replication. Many TLS polymerases contain a UIM that enhances the interaction between monoubiquitylated PCNA and these polymerases, facilitating TLS. PCNA polyubiquitylation is necessary for an error-free DNA-damage-tolerance pathway. SUMOylation of PCNA (predominantly on Lys 164) attracts the anti-recombinogenic helicase Srs2 to inhibit unwanted recombination during DNA synthesis. SUMOylation on (predominantly) Lys 127 inhibits interaction with certain PCNA-binding proteins, such as Eco1.



'reset button' model in which binding of Ubc9 to PCNA, followed by PCNA SUMOylation, may 'kick off' PIP-box proteins from PCNA, allowing a fresh rebinding of other PIP-box proteins once PCNA is de-SUMOylated<sup>27</sup>.

Because Lys 164 of PCNA is a target for ubiquitin and SUMO, the two modifiers compete for the substrate. Remarkably, Ubc9 also physically interacts with Rad18 and Rad5 (ref. 28), suggesting that the enzymes are part of a ubiquitin–SUMO switchboard and that PCNA ubiquitylation and SUMOylation can collaborate during replication. It is theoretically also possible, although currently speculative, that individual subunits of the homotrimeric protein may be modified by different modifiers to regulate events in a stepwise fashion<sup>27</sup>.

### SUMO and structural changes

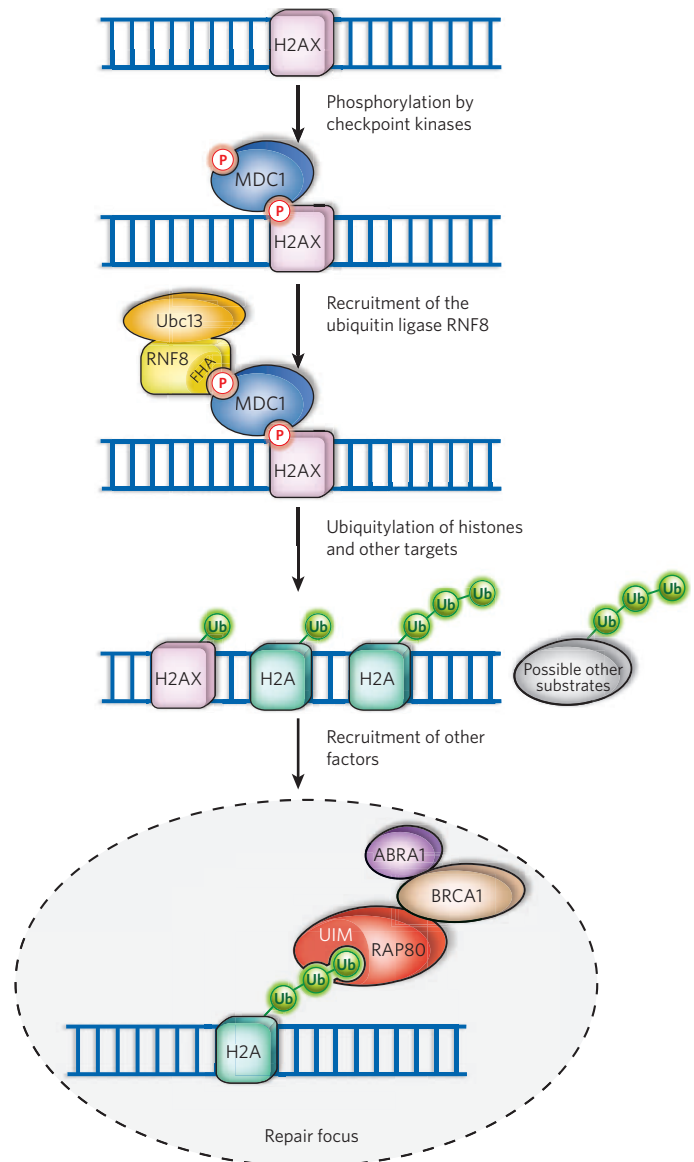
In the cases described above, ubiquitin and SUMO can either attract a protein (TLS polymerases, Srs2) or repel a binding factor (Eco1). An example of how SUMO can directly influence its substrate is found within the base-excision-repair (BER) pathway. BER selectively replaces DNA damage that occurs, usually on its bases, by alkylation, oxidation or deamination. The lesion is recognized by glycosylases, which excise the affected base, resulting in an abasic site (apurinic or apyrimidinic, AP). The phosphodiester bond 5' of the AP site is then cleaved by AP-endonucleases (APE), and the gap is filled in by DNA synthesis and ligation. Thymine-DNA glycosylase (TDG) hydrolyses the N-glycosidic bonds of thymine or uracil when mismatched with guanine<sup>38</sup>. Modification of TDG by SUMO-1 or SUMO-3 causes the dissociation of the enzyme from the product AP site, which results in efficient enzymatic turnover<sup>39</sup>. Controlled dissociation of TDG from the substrate is apparently needed to couple base excision faithfully with APE activity, reducing the exposure of the potentially hazardous AP site. Crystal-structure data of the central domain of the enzyme conjugated to SUMO suggest that conjugated SUMO interacts with a SUMO-binding element in the enzyme. This intramolecular interaction seems to lead to the protrusion of a helix, which might in turn interfere with DNA binding<sup>40</sup>.

### SUMO and nuclear territories

The repair of double-strand breaks (DSBs) provides another example of the way SUMOylation can affect protein function and how it might cooperate with ubiquitylation. DSBs are mostly caused by exogenous agents such as  $\gamma$ -irradiation or certain chemical compounds, but also occasionally occur during faulty replication. They are also actively induced in specific processes, such as meiosis, recombination of immunoglobulin genes and yeast mating-type switching. Major pathways for DSB repair are homologous recombination (HR) and non-homologous end-joining (NHEJ). HR starts with the recruitment of the MRN complex (MRX in yeast) to the break; the complex then tethers the broken ends together and mediates DNA-strand resection, forming single-stranded DNA at the break. The single-stranded DNA is then rapidly coated and protected by the heterotrimeric RPA complex. An important mediator of recombination is Rad52 (and in humans also BRCA2). This wheel-shaped, homo-oligomeric protein removes RPA from chromatin and replaces it with the recombinase Rad51, forming nucleoprotein filaments. Rad51 finally catalyses recombination between homologous sequences<sup>41</sup>.

Rad52 is SUMOylated when the DSB is induced, and is stimulated by the MRX complex<sup>42</sup>. Interestingly, Rad52 SUMOylation rises with the amount of HR substrates, and mutant Rad52 proteins lacking the SUMOylation sites are degraded more rapidly by the ubiquitin–proteasome pathway under these conditions. As Rad52 is stable in the absence of DSBs, this suggests that only recombination-engaged Rad52 proteins are prone to ubiquitin-dependent degradation<sup>42</sup>. SUMOylation may shelter the protein from accelerated degradation, perhaps by sequestering the protein into repair foci.

Undesired recombination between homologous sequences, such as the DNA repeats encoding ribosomal RNA (rDNA), is a particular challenge for HR. In the case of rDNA, this seems to be accomplished in part by restraining the activity of HR within the nucleolus, where the rDNA

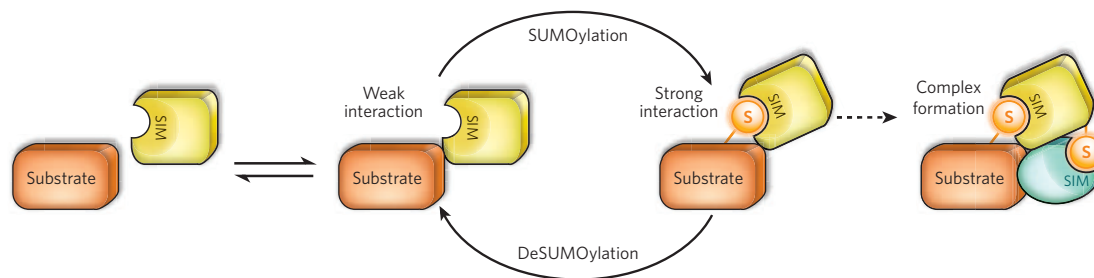


**Figure 4 | Phosphorylation-dependent RNF8 recruitment.** Upon DNA damage, the checkpoint kinase ATM (or ATR) phosphorylates H2AX, recruiting MDC1 to the damage site. The phosphorylation of MDC1 (again by ATM) leads to the phospho-specific binding to the FHA domain of the RING-finger ubiquitin ligase RNF8. RNF8, in conjunction with Ubc13, ubiquitylates targets surrounding the break, including H2A and its variants. RAP80 binds to ubiquitylated targets through a UIM, which leads to the recruitment of the ubiquitin ligase BRCA1 to repair foci (only some proteins of the foci are shown). Interaction between poly-ubiquitylated histones and RAP80 is hypothetical, however. RAP80 has a preference for Lys-63-linked polyubiquitin chains.

usually resides. If DSBs arise at the rDNA locus, they are repaired outside the nucleolus but within the nucleoplasm, as indicated by the formation of extranucleolar repair foci<sup>43</sup>. Indeed, nucleoli seem to be dynamic, as DNA frequently exits and re-enters the nucleolus. In the absence of Rad52 SUMOylation, hyper-recombination within the rDNA cluster occurs, accompanied by the appearance of extragenomic rDNA circles (ERCs)<sup>43</sup>. Whether Rad52 SUMOylation actively prevents productive HR protein assembly in the nucleolus, or whether it immobilizes repair complexes in the nucleoplasm, is an open question.

### Crosstalk between ubiquitylation and phosphorylation

The repair of DSBs in mammalian cells is linked to the *BRCA1* gene, which is mutated in some patients with breast cancer. *BRCA1* encodes



**Figure 5 | Hypothetical model for SUMO-driven protein-complex formation.** Pre-existing weak protein–protein interactions can be strengthened by SUMOylation if the partners bear binding sites for SUMO (SIM). If several proteins of an assembly have SIMs,

SUMOylation can drive multimeric protein-complex (for example, repair foci) formation. DeSUMOylation can trigger disassembly of the complex. Monoubiquitylation and deubiquitylation may have similar roles.

an E3 ubiquitin ligase<sup>44</sup>. Another E3 ligase, RNF8, has also been mapped to damaged sites. RNF8 has an N-terminal FHA domain, which upon DNA damage binds to a phosphorylated region of MDC1. This in turn binds phosphorylated H2AX, a well-known chromatin mark for DNA damage. The accumulation of RNF8 at DNA lesions results in the recruitment of 53BP1 and BRCA1, causing extensive ubiquitylation of proteins at the damage site<sup>45–47</sup>. Known targets of RNF8 are the histone H2A and its variant H2AX<sup>45,47</sup>, of which the former has previously been identified as an ultraviolet-light-induced ubiquitylation substrate<sup>48</sup>. RNF8 collaborates with the E2 enzyme UBC13, which also functions in the error-free branch of the *RAD6* pathway. Indeed, in UBC13-deficient chicken cells, BRCA1 does not accumulate at damaged sites, and H2AX ubiquitylation does not occur after irradiation<sup>49</sup>.

The protein RAP80, which contains a UIM, is necessary for BRCA1 accumulation at damaged sites<sup>50–52</sup>. In UBC13- and RNF8-depleted cells, RAP80 accumulation is abolished<sup>45</sup>. Because UBC13 is involved in the catalysis of Lys-63-linked polyubiquitin chains, and because RAP80 exhibits preference for this type of modification, it is likely that ubiquitin has a non-proteolytic role in this pathway. It is unclear whether RAP80 binds directly to (poly)ubiquitylated H2A, a modified histone variant, or to an unidentified RNF8 target (Fig. 4). As well as facilitating the recruitment of downstream factors, histone ubiquitylation might affect chromatin structure, possibly leading to repair or damage signalling.

A similar accumulation of ubiquitin has been seen at sites of NER (refs 48, 53). Like H2AX phosphorylation<sup>54</sup>, H2A ubiquitylation also depends on DNA damage<sup>48</sup>. However, whether different types of DNA lesions lead to a similar DDR response is unclear.

### Regulating DDR signalling

Ubiquitylation and SUMOylation are transient and reversible modifications, and the enzymes that remove the modifiers have an important regulatory role. An example of the importance of deubiquitylation is given in the Fanconi anaemia pathway, which mediates the resolution of DNA interstrand crosslinks. Damage of this type is highly cytotoxic and needs a combination of the HR, NER and TLS pathways. Deficiency in one of the known 13 FA genes results in hypersensitivity to DNA crosslinkers. Most of the identified Fanconi anaemia proteins constitute a single, large, multi-subunit E3 ubiquitin ligase, and two others, FANCD2 and FANCI, are the corresponding substrates, which relocate to chromatin when ubiquitylated<sup>55–57</sup>. The function of these proteins at the chromatin, and in the subsequent steps, is currently unknown, however.

The deubiquitylation enzyme USP1 acts on both FANCD2 and FANCI and also on PCNA<sup>57–59</sup>. Loss of USP1 activity results in the ubiquitylation of these proteins, even if there is no DNA damage<sup>57–59</sup>. Chicken cells in which the gene encoding USP1 has been knocked out display hypersensitivity to DNA crosslinkers, the hallmark of Fanconi anaemia<sup>60</sup>, suggesting that the deubiquitylation of FANCD2 and FANCI is crucial for repair. USP1 contains an internal domain related to ubiquitin's C terminus, which promotes autocleavage<sup>58</sup>. However, because the cleaved enzyme is still proteolytically active<sup>61</sup> and an engineered non-cleavable

variant can partly complement USP1-deficient cells<sup>60</sup>, the function of USP1 autocleavage is unknown. Nevertheless, because both the ubiquitylation and the deubiquitylation of FANCD2 and FANCI seem to be crucial for normal DNA-crosslink repair, it is likely that different levels of regulation are important to guide and coordinate repair enzymes at the sites of lesions.

### Thoughts and ideas

Future studies will no doubt reveal further links between ubiquitylation and SUMOylation and human diseases, including DNA-repair disorders. Both types of modification function in two general ways: as an adhesive or a repellent. The various surfaces of the two modifiers provide binding sites for protein–protein interactions, but their bulkiness can block other interactions. Several interactions of ubiquitin or SUMO with cognate binding proteins are relatively weak. Moreover, sometimes binding occurs even in the absence of modification, but the attached modifiers greatly strengthen the associations. It therefore seems plausible that protein assemblies may be firmly held in place by additive weak interactions between several modifiers and several modifier-binding modules present on different proteins of the assembly. As a result, the assembly of complexes is less efficient when there is no damage; that is, when proteins are unmodified (Fig. 5). In this model, disassembly is initiated by deubiquitylation and deSUMOylation enzymes or, alternatively, by the repellent activities of the modifiers. Because many nuclear functions take place within large protein assemblies, it will be interesting to see whether key functions of the two modifiers are perhaps to define nuclear territories, to assemble or disassemble nuclear activities, or to provide directionality to a pathway by a stepwise organization of protein complexes that should function one after the other. ■

1. Kerscher, O., Felberbaum, R. & Hochstrasser, M. Modification of proteins by ubiquitin and ubiquitin-like proteins. *Annu. Rev. Cell Dev. Biol.* **22**, 159–180 (2006).
2. Somesh, B. P. *et al.* Communication between distant sites in RNA polymerase II through ubiquitylation factors and the polymerase CTD. *Cell* **129**, 57–68 (2007).
3. Tatham, M. H. *et al.* RNF4 is a poly-SUMO-specific E3 ubiquitin ligase required for arsenic-induced PML degradation. *Nature Cell Biol.* **10**, 538–546 (2008).
4. Peng, J. *et al.* A proteomics approach to understanding protein ubiquitination. *Nature Biotechnol.* **21**, 921–926 (2003).
5. Kirkin, V. & Dikic, I. Role of ubiquitin- and Ubl-binding proteins in cell signaling. *Curr. Opin. Cell Biol.* **19**, 199–205 (2007).
6. Hecker, C. M., Rabiller, M., Haglund, K., Bayer, P. & Dikic, I. Specification of SUMO1- and SUMO2-interacting motifs. *J. Biol. Chem.* **281**, 16117–16127 (2006).
7. Varadan, R., Assfalg, M., Raasi, S., Pickart, C. & Fushman, D. Structural determinants for selective recognition of a Lys48-linked polyubiquitin chain by a UBA domain. *Mol. Cell* **18**, 687–698 (2005).
8. Panse, V. G., Hardeland, U., Werner, T., Kuster, B. & Hurt, E. A proteome-wide approach identifies sumoylated substrate proteins in yeast. *J. Biol. Chem.* **279**, 41346–41351 (2004).
9. Hoeijmakers, J. H. Genome maintenance mechanisms for preventing cancer. *Nature* **411**, 366–374 (2001).
10. Branzel, D. & Foiani, M. Regulation of DNA repair throughout the cell cycle. *Nature Rev. Mol. Cell Biol.* **9**, 297–308 (2008).
11. Harrison, J. C. & Haber, J. E. Surviving the breakup: the DNA damage checkpoint. *Annu. Rev. Genet.* **40**, 209–235 (2006).
12. Woudstra, E. C. *et al.* A Rad26–Def1 complex coordinates repair and RNA pol II proteolysis in response to DNA damage. *Nature* **415**, 929–933 (2002).
13. Min, J. H. & Pavletich, N. P. Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature* **449**, 570–575 (2007).



14. Chu, G. & Chang, E. Xeroderma pigmentosum group E cells lack a nuclear factor that binds to damaged DNA. *Science* **242**, 564–567 (1988).
15. Araki, M. *et al.* Centrosome protein centrin 2/caltractin 1 is part of the xeroderma pigmentosum group C complex that initiates global genome nucleotide excision repair. *J. Biol. Chem.* **276**, 18665–18672 (2001).
16. Masutani, C. *et al.* Purification and cloning of a nucleotide excision repair complex involving the xeroderma pigmentosum group C protein and a human homologue of yeast RAD23. *EMBO J.* **13**, 1831–1843 (1994).
17. Verma, R., Oania, R., Graumann, J. & Deshaies, R. J. Multiubiquitin chain receptors define a layer of substrate selectivity in the ubiquitin–proteasome system. *Cell* **118**, 99–110 (2004).
18. Gillette, T. G. *et al.* Distinct functions of the ubiquitin–proteasome pathway influence nucleotide excision repair. *EMBO J.* **25**, 2529–2538 (2006).
19. Lommel, L., Ortolan, T., Chen, L., Madura, K. & Sweder, K. S. Proteolysis of a nucleotide excision repair protein by the 26 S proteasome. *Curr. Genet.* **42**, 9–20 (2002).
20. Ng, J. M. *et al.* A novel regulation mechanism of DNA repair by damage-induced and RAD23-dependent stabilization of xeroderma pigmentosum group C protein. *Genes Dev.* **17**, 1630–1645 (2003).
21. Sugawara, K. *et al.* UV-induced ubiquitylation of XPC protein mediated by UV-DDB–ubiquitin ligase complex. *Cell* **121**, 387–400 (2005).
22. Groisman, R. *et al.* The ubiquitin ligase activity in the DDB2 and CSA complexes is differentially regulated by the COP9 signalosome in response to DNA damage. *Cell* **113**, 357–367 (2003).
23. Chen, X., Zhang, Y., Douglas, L. & Zhou, P. UV-damaged DNA-binding proteins are targets of CUL4A-mediated ubiquitination and degradation. *J. Biol. Chem.* **276**, 48175–48182 (2001).
24. Russell, S. J., Reed, S. H., Huang, W., Friedberg, E. C. & Johnston, S. A. The 19S regulatory complex of the proteasome functions independently of proteolysis in nucleotide excision repair. *Mol. Cell* **3**, 687–695 (1999).
25. Smith, D. M., Benaroudj, N. & Goldberg, A. Proteasomes and their associated ATPases: a destructive combination. *J. Struct. Biol.* **156**, 72–83 (2006).
26. Jentsch, S. & Rumpf, S. Cdc48 (p97): a 'molecular gearbox' in the ubiquitin pathway? *Trends Biochem. Sci.* **32**, 6–11 (2007).
27. Moldovan, G. L., Pfander, B. & Jentsch, S. PCNA, the maestro of the replication fork. *Cell* **129**, 665–679 (2007).
28. Hoegge, C., Pfander, B., Moldovan, G. L., Pyrowolakis, G. & Jentsch, S. RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature* **419**, 135–141 (2002).
- This paper determines that PCNA is the key target that steers RAD6-dependent repair.**
29. Ulrich, H. D. & Jentsch, S. Two RING finger proteins mediate cooperation between ubiquitin-conjugating enzymes in DNA repair. *EMBO J.* **19**, 3388–3397 (2000).
30. Jentsch, S., McGrath, J. P. & Varshavsky, A. The yeast DNA repair gene RAD6 encodes a ubiquitin-conjugating enzyme. *Nature* **329**, 131–134 (1987).
31. Davies, A. A., Huttner, D., Daigaku, Y., Chen, S. & Ulrich, H. D. Activation of ubiquitin-dependent DNA damage bypass is mediated by replication protein A. *Mol. Cell* **29**, 625–636 (2008).
32. Friedberg, E. C., Lehmann, A. R. & Fuchs, R. P. Trading places: how do DNA polymerases switch during translesion DNA synthesis? *Mol. Cell* **18**, 499–505 (2005).
33. Biemko, M. *et al.* Ubiquitin-binding domains in Y-family polymerases regulate translesion synthesis. *Science* **310**, 1821–1824 (2005).
34. Kannouche, P. L., Wing, J. & Lehmann, A. R. Interaction of human DNA polymerase  $\eta$  with monoubiquitinated PCNA: a possible mechanism for the polymerase switch in response to DNA damage. *Mol. Cell* **14**, 491–500 (2004).
35. Papouli, E. *et al.* Crosstalk between SUMO and ubiquitin on PCNA is mediated by recruitment of the helicase Srs2. *Mol. Cell* **19**, 123–133 (2005).
36. Pfander, B., Moldovan, G. L., Sacher, M., Hoegge, C. & Jentsch, S. SUMO-modified PCNA recruits Srs2 to prevent recombination during S phase. *Nature* **436**, 428–433 (2005).
- References 35 and 36 show that the helicase Srs2 binds to SUMOylated PCNA, preventing unwanted recombination at the replication fork.**
37. Moldovan, G. L., Pfander, B. & Jentsch, S. PCNA controls establishment of sister chromatid cohesion during S phase. *Mol. Cell* **23**, 723–732 (2006).
- This is the first example of a possible repellent function of SUMO.**
38. Lindahl, T., Karran, P. & Wood, R. D. DNA excision repair pathways. *Curr. Opin. Genet. Dev.* **7**, 158–169 (1997).
39. Hardeland, U., Steinacher, R., Jiricny, J. & Schar, P. Modification of the human thymine-DNA glycosylase by ubiquitin-like proteins facilitates enzymatic turnover. *EMBO J.* **21**, 1456–1464 (2002).
- This paper shows that SUMOylation can regulate enzyme activity.**
40. Baba, D. *et al.* Crystal structure of thymine DNA glycosylase conjugated to SUMO-1. *Nature* **435**, 979–982 (2005).
41. van Gent, D. C., Hoeijmakers, J. H. & Kanaar, R. Chromosomal stability and the DNA double-stranded break connection. *Nature Rev. Genet.* **2**, 196–206 (2001).
42. Sacher, M., Pfander, B., Hoegge, C. & Jentsch, S. Control of Rad52 recombination activity by double-strand break-induced SUMO modification. *Nature Cell Biol.* **8**, 1284–1290 (2006).
43. Torres-Rosell, J. *et al.* The Smc5–Smc6 complex and SUMO modification of Rad52 regulates recombinational repair at the ribosomal gene locus. *Nature Cell Biol.* **9**, 923–931 (2007).
- This paper shows that SUMOylation of Rad52 alters the localization of repair foci.**
44. Lorick, K. L. *et al.* RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination. *Proc. Natl Acad. Sci. USA* **96**, 11364–11369 (1999).
45. Huen, M. S. *et al.* RNF8 transduces the DNA-damage signal via histone ubiquitylation and checkpoint protein assembly. *Cell* **131**, 901–914 (2007).
46. Kolas, N. K. *et al.* Orchestration of the DNA-damage response by the RNF8 ubiquitin ligase. *Science* **318**, 1637–1640 (2007).
47. Mailand, N. *et al.* RNF8 ubiquitylates histones at DNA double-strand breaks and promotes assembly of repair proteins. *Cell* **131**, 887–900 (2007).
- References 45–47 show how phosphorylation and ubiquitylation within the DDR are linked by the RNF8 E3 ligase.**
48. Bergink, S. *et al.* DNA damage triggers nucleotide excision repair-dependent monoubiquitylation of histone H2A. *Genes Dev.* **20**, 1343–1352 (2006).
49. Zhao, G. Y. *et al.* A critical role for the ubiquitin-conjugating enzyme Ubc13 in initiating homologous recombination. *Mol. Cell* **25**, 663–675 (2007).
50. Kim, H., Chen, J. & Yu, X. Ubiquitin-binding protein RAP80 mediates BRCA1-dependent DNA damage response. *Science* **316**, 1202–1205 (2007).
51. Sobhian, B. *et al.* RAP80 targets BRCA1 to specific ubiquitin structures at DNA damage sites. *Science* **316**, 1198–1202 (2007).
52. Wang, B. *et al.* Abraxas and RAP80 form a BRCA1 protein complex required for the DNA damage response. *Science* **316**, 1194–1198 (2007).
53. Wang, H. *et al.* Histone H3 and H4 ubiquitylation by the CUL4–DDB–ROC1 ubiquitin ligase facilitates cellular response to DNA damage. *Mol. Cell* **22**, 383–394 (2006).
54. O'Driscoll, M., Ruiz-Perez, V. L., Woods, C. G., Jeggo, P. A. & Goodship, J. A. A splicing mutation affecting expression of ataxia-telangiectasia and Rad3-related protein (ATR) results in Seckel syndrome. *Nature Genet.* **33**, 497–501 (2003).
55. Garcia-Higuera, I. *et al.* Interaction of the Fanconi anemia proteins and BRCA1 in a common pathway. *Mol. Cell* **7**, 249–262 (2001).
56. Meetei, A. R. *et al.* A novel ubiquitin ligase is deficient in Fanconi anemia. *Nature Genet.* **35**, 165–170 (2003).
57. Smogorzewska, A. *et al.* Identification of the FANCI protein, a monoubiquitinated FANCD2 paralog required for DNA repair. *Cell* **129**, 289–301 (2007).
58. Huang, T. T. *et al.* Regulation of monoubiquitinated PCNA by DUB autocleavage. *Nature Cell Biol.* **8**, 341–347 (2006).
59. Nijman, S. M. *et al.* The deubiquitinating enzyme USP1 regulates the Fanconi anemia pathway. *Mol. Cell* **17**, 331–339 (2005).
60. Oestergaard, V. H. *et al.* Deubiquitination of FANCD2 is required for DNA crosslink repair. *Mol. Cell* **28**, 798–809 (2007).
- References 59 and 60 show the importance of deubiquitylation in DNA repair.**
61. Cohn, M. A. *et al.* A UAF1-containing multisubunit protein complex regulates the Fanconi anemia pathway. *Mol. Cell* **28**, 786–797 (2007).
62. Arakawa, H. *et al.* A role for PCNA ubiquitination in immunoglobulin hypermutation. *PLoS Biol.* **4**, e366 (2006).
63. Frampton, J. *et al.* Postreplication repair and PCNA modification in *Schizosaccharomyces pombe*. *Mol. Biol. Cell* **17**, 2976–2985 (2006).
64. Leach, C. A. & Michael, W. M. Ubiquitin/SUMO modification of PCNA promotes replication fork progression in *Xenopus laevis* egg extracts. *J. Cell Biol.* **171**, 947–954 (2005).
65. Fu, Y. *et al.* Rad6–Rad18 mediates a eukaryotic SOS response by ubiquitinating the 9–1–1 checkpoint clamp. *Cell* **133**, 601–611 (2008).
66. Wang, Q. E. *et al.* DNA repair factor XPC is modified by SUMO-1 and ubiquitin following UV irradiation. *Nucleic Acids Res.* **33**, 4023–4034 (2005).
67. Trickey, M., Grimaldi, M. & Yamano, H. The anaphase-promoting complex/cyclosome controls repair and recombination by ubiquitylating Rhp54 in fission yeast. *Mol. Cell. Biol.* **28**, 3905–3916 (2008).
68. Yurchenko, V., Xue, Z. & Sadofsky, M. J. SUMO modification of human XRCC4 regulates its localization and function in DNA double-strand break repair. *Mol. Cell. Biol.* **26**, 1786–1794 (2006).
69. Zhao, X. & Blobel, G. A SUMO ligase is part of a nuclear multiprotein complex that affects DNA repair and chromosomal organization. *Proc. Natl Acad. Sci. USA* **102**, 4777–4782 (2005).
70. Postow, L. *et al.* Ku80 removal from DNA through double strand break-induced ubiquitylation. *J. Cell Biol.* **182**, 467–479 (2008).
71. Eladad, S. *et al.* Intra-nuclear trafficking of the BLM helicase to DNA damage-induced foci is regulated by SUMO modification. *Hum. Mol. Genet.* **14**, 1351–1365 (2005).
72. Kawabe, Y. *et al.* Covalent modification of the Werner's syndrome gene product with the ubiquitin-related protein, SUMO-1. *J. Biol. Chem.* **275**, 20963–20966 (2000).
73. Branzei, D. *et al.* Ubc9- and Mms21-mediated sumoylation counteracts recombinogenic events at damaged replication forks. *Cell* **127**, 509–522 (2006).
74. Mao, Y., Desai, S. D. & Liu, L. F. SUMO-1 conjugation to human DNA topoisomerase II isozymes. *J. Biol. Chem.* **275**, 26066–26073 (2000).
75. Bachant, J., Alcasabas, A., Blat, Y., Kleckner, N. & Elledge, S. J. The SUMO-1 isopeptidase Smt4 is linked to centromeric cohesion through SUMO-1 modification of DNA topoisomerase II. *Mol. Cell* **9**, 1169–1182 (2002).

**Acknowledgements** We thank A. Buchberger, S. Müller and members of the Department of Molecular Cell Biology at the Max Planck Institute for Biochemistry for discussions. S.J. is supported by the Max Planck Society, the Centre for Integrated Protein Science Munich, the Deutsche Krebshilfe and the RUBICON Network of Excellence.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Correspondence should be addressed to S.J. (Jentsch@biochem.mpg.de).

# The oldest articulated osteichthyan reveals mosaic gnathostome characters

Min Zhu<sup>1</sup>, Wenjin Zhao<sup>1</sup>, Liantao Jia<sup>1</sup>, Jing Lu<sup>1,2</sup>, Tuo Qiao<sup>1,2</sup> & Qingming Qu<sup>1,2</sup>

**The evolutionary history of osteichthyans (bony fishes plus tetrapods) extends back to the Ludlow epoch of the Silurian period. However, these Silurian forms have been documented exclusively by fragmentary fossils. Here we report the discovery of an exceptionally preserved primitive fish from the Ludlow of Yunnan, China, that represents the oldest near-complete gnathostome (jawed vertebrate). The postcranial skeleton of this fish includes a primitive pectoral girdle and median fin spine as in non-osteichthyan gnathostomes, but a derived macromeric squamation as in crown osteichthyans, and substantiates the unexpected mix of postcranial features in basal sarcopterygians, previously restored from the disarticulated remains of *Psarolepis*. As the oldest articulated sarcopterygian, the new taxon offers insights into the origin and early divergence of osteichthyans, and indicates that the minimum date for the actinopterygian–sarcopterygian split was no later than 419 million years ago.**

Osteichthyans, which fall into two major monophyletic groups, namely actinopterygians (bichirs, sturgeons, gars, bowfins and teleosts) and sarcopterygians (coelacanth, lungfishes and tetrapods), make up 98% of recognized living vertebrate species<sup>1</sup>. The rise of osteichthyans from other primitive gnathostomes is a key transition in vertebrate evolution, yet this transition is poorly documented by the fossil discoveries of the last century<sup>2</sup>. Major morphological gaps existed between actinopterygians and sarcopterygians, and between osteichthyans and non-osteichthyan gnathostomes (chondrichthyans, acanthodians and placoderms). The past decade has seen the gap narrowing with fossil discoveries such as *Psarolepis*<sup>3–5</sup>, *Ligulalepis*<sup>6,7</sup> and *Dialipina*<sup>8</sup>, which show fascinating combinations of osteichthyan and non-osteichthyan gnathostome characters, providing new opportunities for studying the polarity and evolution of these characters. However, the basal osteichthyan phylogeny remains uncertain<sup>4,6,9–11</sup> owing to the large number of unknown character states in these early forms and the provisional assignment of disarticulated remains to a single taxon. A better understanding of these fishes is therefore crucial in reconstructing the part of phylogeny close to the split between actinopterygians and sarcopterygians.

The fossil record of osteichthyans extends back to the Ludlow epoch of the Silurian period, ~422 million years (Myr) ago<sup>2,11</sup>, yet is documented by fragmentary fossils<sup>12–17</sup>. *Dialipina salgueiroensis* from the Emsian of the Canadian Arctic<sup>8</sup> (~400 Myr ago, *serotinus* conodont zone<sup>18</sup>), referred to the actinopterygians<sup>2,8,9,11,19</sup> or stem osteichthyans<sup>10</sup>, represents the oldest articulated and complete osteichthyan known thus far. Two articulated stem lungfishes, *Porolepis* from Spitsbergen<sup>20</sup> and *Uranolophus* from Wyoming<sup>21</sup>, are older (Pragian, ~411 Myr ago) but less complete than *D. salgueiroensis*. *Psarolepis* from the latest Pridoli to the Lochkovian of China and Vietnam<sup>3–5,17,22</sup> is another early osteichthyan with the postcranial condition partially known, albeit based on disarticulated remains. Its tentative reconstruction<sup>4</sup> shows an enigmatic fish bearing mosaic gnathostome features, but the association of parts from different individuals requires testing against articulated specimens.

The new articulated fish from the Silurian of China closely resembles *Psarolepis* in cranial features, although it bears ridged ornamentation,

as present in basal actinopterygians, rather than the cosmine of early sarcopterygians<sup>9</sup>. This discovery extends the earliest record of near-complete articulated osteichthyans to at least 8 Myr earlier in geological history, and significantly enhances our growing knowledge of the origin of osteichthyan morphology.

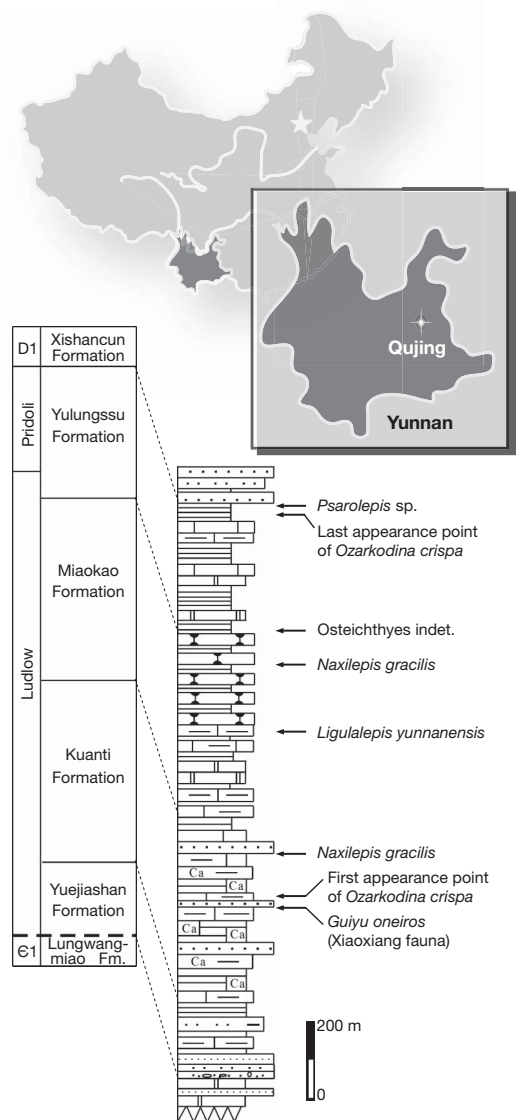
## Geological framework

The new fish material was recovered from the muddy limestone of the Kuanti Formation immediately beneath the first appearance point of *Ozarkodina crista*<sup>23</sup> at a locality near Xiaoxiang Reservoir, Qujing, Yunnan, China (Fig. 1). The marine Silurian strata in Qujing are subdivided into four formations in ascending order: the Yuejiashan, Kuanti, Miaokao and Yulongssu formations<sup>24–26</sup>. Early fishes are recorded from the sequence in association with rich invertebrates such as corals, brachiopods, cephalopods, ostracods, bryozoans and trilobites<sup>25</sup>, and include *Psarolepis*, an indeterminate osteichthyan from the Yulongssu Formation<sup>15,17</sup>, and two 'actinopterygians' *Naxilepis gracilis* and *Ligulalepis yunnanensis* from the Miaokao and Kuanti formations<sup>16</sup>. The latter three forms are only represented by scales, and the assignment of *Naxilepis* and *L. yunnanensis* to the actinopterygians is doubtful due to the meagreness of available data. The new fish is found from the bed ~140 m below the lowest occurrence of *Naxilepis*, thereby representing the earliest record of osteichthyans in China. The fish fauna (here named the Xiaoxiang fauna, characterized by the early diversification of gnathostomes) from the site also includes the agnathans, placoderms and acanthodians under study. The late Ludlow (Silurian) age of the new fish bed is mainly derived from the conodont zonation<sup>27</sup>, although the earlier age assessment (early Ludlow or late Wenlock) of the Kuanti Formation had been proposed based on invertebrate or vertebrate assemblages<sup>25,26,28</sup>.

Osteichthyes Huxley, 1880  
Sarcopterygii (Romer, 1955)  
*Guiyu oneiros* gen. et sp. nov.

**Etymology.** The generic name derives from *gui* (Chinese Pinyin), meaning ghost or secret, and *yu* (Chinese Pinyin), meaning fish.

<sup>1</sup>Key Laboratory of Evolutionary Systematics of Vertebrates, Institute of Vertebrate Paleontology and Paleoanthropology (IVPP), Chinese Academy of Sciences, PO Box 643, Beijing 100044, China. <sup>2</sup>Graduate School, Chinese Academy of Sciences, Beijing 100039, China.



**Figure 1 | Summary Silurian sequence in Qujing (Yunnan, China), showing the stratigraphic positions of *Guiyu oneiros* gen. et sp. nov. and other bony fishes.** The stratigraphic data are taken from ref. 25.

The specific name is from the Greek *oneiros*, dream, alluding to a dream fish with mosaic gnathostome characters.

**Holotype.** IVPP (Institute of Vertebrate Paleontology and Paleoanthropology, Beijing) V15541, a near-complete fish in part and counterpart.

**Referred material.** Skulls, V15542.1-6; lower jaws, V15542.7-13; scales, V15542.14-16.

**Horizon and locality.** Late Ludlow, Silurian, Kuantu Formation; Qujing, Yunnan, China.

**Diagnosis.** *Guiyu* differs from *Psarolepis*<sup>3-5,17</sup> and *Achoania*<sup>29,30</sup> in its ganoine sculpturing of dermal skeleton, anterior nostril without dermintermedial process, no teeth on median rostral, premaxillary without inturned symphyseal process, presence of prerostral bone, dentary teeth reaching anterior end of lower jaw, and lacking infradentary foramina.

## Description

The holotype in its part and counterpart represents an articulated and three-dimensionally preserved bony fish lacking only the caudal fin (Figs 2a, b and 3a, b). The dermal bones and large rhomboid scales are ornamented with ganoine sculpturing of coarse, irregular or patterned

longitudinal ridges (Supplementary Information 2, section 9), characteristic of lower actinopterygians<sup>8,19,31-34</sup> and *Ligulalepis*<sup>6,7,35</sup>. The ganoine-like tissue is also present in the surface covering of acanthodians and sarcopterygians, suggesting that ganoine might be plesiomorphic for crown osteichthyans<sup>36</sup>. The fish has a preserved length of ~26 cm and a preserved maximum depth (although laterally compressed and distorted) of ~11 cm. The head occupies about 23% of the total body length (to the base of the caudal fin), and the body is about 2.5 times as long as it is deep, giving *Guiyu* a fusiform streamlined appearance.

The skull roof is divided into the parietal and postparietal shields by the dermal intracranial joint (Fig. 4a). The parietal shield is about as wide as it is long, and has the supraorbital sensory canal medial to the anterodorsally facing anterior nostril. The dermal bone pattern is more evidently shown in several isolated skull specimens (for example, V15542.1, Fig. 4b). Unlike *Psarolepis* and *Achoania*, the large median rostral does not form the outer dental arcade, and the premaxillary lacks the obvious inturned symphyseal process in anterior view (Fig. 4c). A separate triangular bone (preorbital, Fig. 4c) posterodorsal to the premaxillary, rather than the posterodorsal process of the premaxillary, encloses the posterior nostril; this resembles *Onychodus*<sup>37</sup> and possibly *Achoania*<sup>29</sup> but differs from *Psarolepis*<sup>3</sup>. The 'lacrimal' in *Onychodus*<sup>37</sup>, which has a more anterior position than the lacrimal of rhipidistians<sup>2</sup>, might be homologous with the preorbital of actinopterygians and *Guiyu*. Noteworthy is an unpaired tooth-bearing bone behind the premaxillaries in ventral view (Fig. 4e, f). This bone has never been found in known osteichthyans, yet topologically corresponds to the prerostral or premedian plate of placoderms<sup>2,38</sup>. The postparietal shield is as long as the parietal shield, as in some later sarcopterygians<sup>38</sup>. In *Psarolepis*, the parietal shield is about 1.5 times the postparietal shield in length. The middle and posterior pit-lines lie close to the midline of the shield, and the otic canal runs through the supratemporal and tabular, and connects the supratemporal commissural canal in the lateral extrascapular (Fig. 2b). Anterior to the extratemporal—which is a small bone lateral to the tabular and the anterior half of the lateral extrascapular—an elongate bone between the supratemporal and the opercular represents an accessory extratemporal. Three extrascapular bones are situated behind the postparietal shield; this resembles onychodonts and rhipidistians but differs from actinopterygians and most coelacanths.

Concerning the neurocranial features, *Guiyu* shows a striking resemblance to *Psarolepis*, *Achoania* and *Onychodus*. They share large pear-shaped internasal cavities, a long and narrow internasal septum in front of a small triangular parasphenoid, and the same position of ethmoid articulation on the postnasal wall (Fig. 4e, f). The neurocranium of *Guiyu* is much less ossified than that of *Psarolepis* and *Achoania*, raising difficulties in locating the anatomical structures on its lateral sides (Supplementary Information 2, section 7). However, the postorbital pillar is distinguishable as a bridge connecting the basipterygoid process and the side of the neurocranium.

The dermal bones of the cheek and operculo-gular system show a similar pattern to those of early actinopterygians<sup>31,32</sup>, *Psarolepis* and onychodonts. Three branchiostegal rays on the left side are preserved below the big opercular and subopercular bones. The large gular bones occupy most of the space between the lower jaws as in *Onychodus*. The lower jaws of the holotype are anteriorly dislocated and broken in their upper half; however, the posterior extremity of the left lower jaw retains its natural position to the upper jaw. As supplemented by some isolated specimens (Fig. 4g, j-m), the lower jaw of *Guiyu* is nearly identical in shape to that of *Psarolepis*, and is anteriorly reflexed. The mandibular sensory canal pierces through four infradentaries, leaving a series of openings on the external surface. Posteriorly, a shallow area devoid of ornamentation represents the part of the lower jaw overlapped by the maxillary, as in early sarcopterygians<sup>30</sup>. In early actinopterygians, this area is usually more extensive<sup>32</sup>. The dentary bears a row of sharp, conical but slender teeth on the dorsal margin



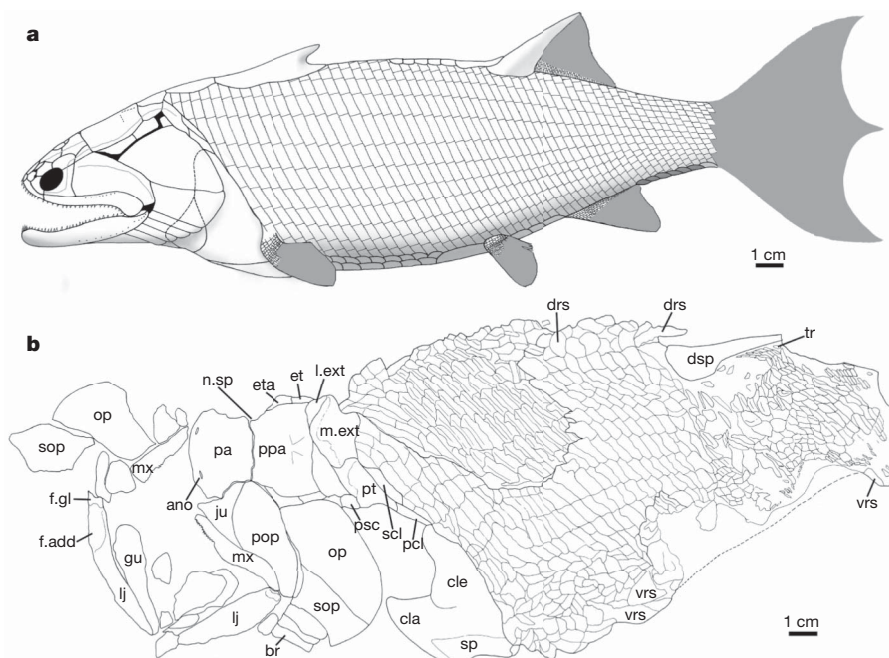


**Figure 2** | *Guiyu oneiros* gen. et sp. nov., holotype V15541. **a, b**, A near-complete fish in part and counterpart. **c**, Close-up view of the anterior portion of the trunk shield in dorsal view, showing MD1 and MD2 flanked by

rhomboid scales. **d**, Close-up view of the dorsal fin spine. MD1, first median dorsal plate; MD2, second median dorsal plate. Scale bar, 1 cm.

interior to a band of smaller tooth-like denticles that are present in *Lophosteus*, *Andreolepis*<sup>11</sup> and many crown osteichthyans. Each marginal tooth has a large pulp cavity extending close to the tip, and lacks the acrodin (Supplementary Information 2, section 9c, e). *Guiyu*

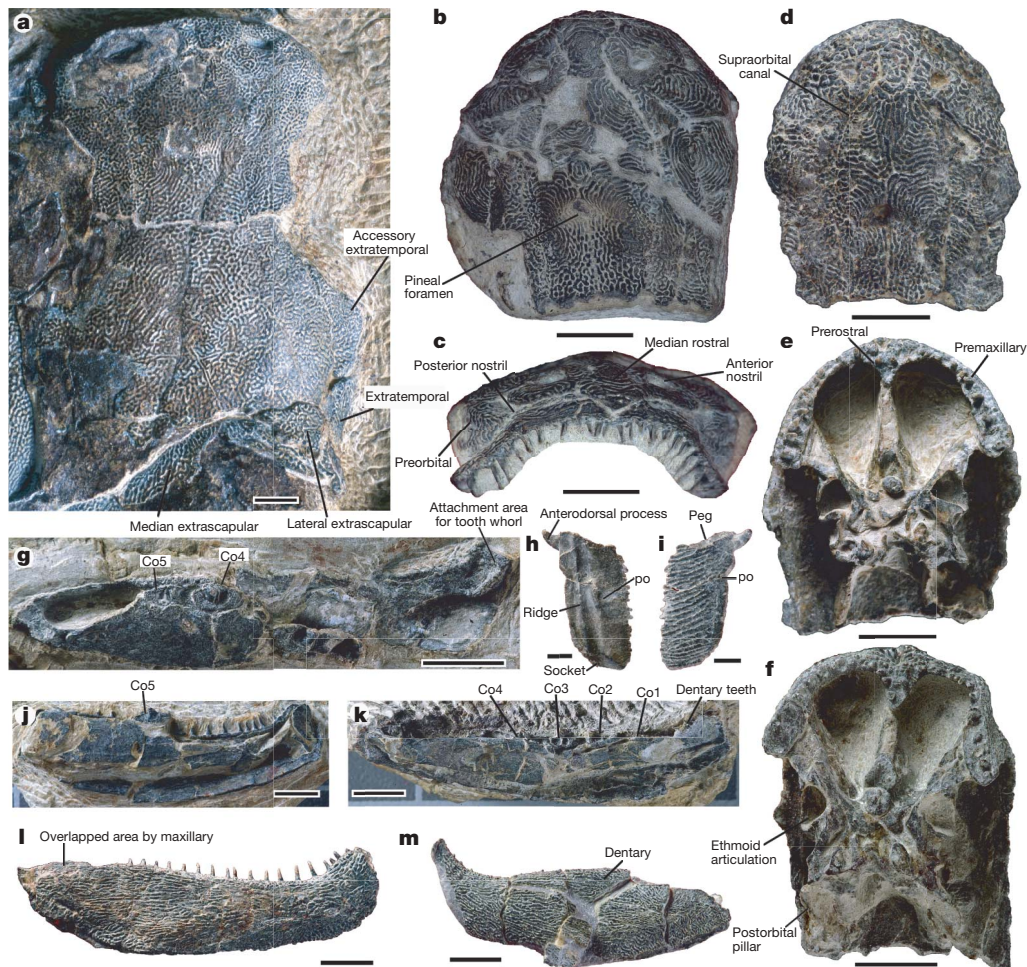
resembles *Psarolepis* in having five coronoids separating the dentary and prearticular bones, a supporting Meckelian lamina for the parasymphyseal tooth whorl (albeit less ossified and developed), and a relatively short adductor fossa. The second to fourth coronoids carry



**Figure 3** | *Guiyu oneiros* gen. et sp. nov. **a**, Restoration of the entire fish in lateral view. **b**, Interpretive drawing of the holotype V15541. Areas shaded in grey are unknown, and are reconstructed from other early osteichthyans<sup>8,37</sup>. ano, anterior nostril; br, branchiostegal ray; cla, clavicle; cle, cleithrum; drs, dorsal ridge scale; dsp, dorsal fin spine; et, extratemporal; eta, accessory extratemporal; f.add, adductor fossa; f.gl, glenoid fossa; gu, gular; ju, jugal;

Lext, lateral extrascapular; lj, lower jaw; m.ext, median extrascapular; mx, maxillary; n.sp., spiracular notch; op, opercular; pa, parietal shield; pcl, postcleithrum; pop, preopercular; ppa, postparietal shield; psc, presupracleithrum; pt, post-temporal; scl, supracleithrum; sop, subopercular; sp., pectoral spine; tr, lepidotrichia; vrs, ventral ridge scale.





**Figure 4** | *Guiyu oneiros* gen. et sp. nov. **a**, Close-up view of the head in dorsal view, holotype V15541. **b–f**, Isolated anterior cranial portions: **b**, **c**, dorsal and anterior views (V15542.1); **d**, **e**, dorsal and ventral views (V15542.2); **f**, ventral view (V15542.3). **g**, A lower jaw in internal view,

V15542.7. **h**, **i**, An isolated scale in internal and external views, V15542.14. **j–m**, Lower jaws: **j**, internal view (V15542.8); **k**, internal view (V15542.9); **l**, external view (V15542.10); **m**, external view (V15542.11). Co1–5, first to fifth coronoid; po, pore of lateral line. Scale bar: 2 mm (**h**, **i**); 5 mm (**a–g**, **j–m**).

large tusks and a narrow lateral band with 2–4 denticle rows. The tusk has a simple histological structure and lacks any in-folding of dentine near its root, as documented by the fractured surfaces (Supplementary Information 2, section 9f, g).

The pectoral girdle of *Guiyu* is remarkable in its separate fin spine (Fig. 2b), as in *Psarolepis*<sup>4,5</sup>, the early chondrichthyan *Doliodus*<sup>39</sup>, placoderms and acanthodians<sup>2</sup>. A small bone (psc, Fig. 3b) in front of the post-temporal corresponds to the presupracleithrum of early actinopterygians<sup>32,33</sup>. The extensive postbranchial lamina is covered with posteriorly pointed tubercles as in *Psarolepis*, actinopterygians and placoderms. Immediately behind the median extrascapular, the postcranial skeleton is conspicuous in having two large median dorsal plates (MD1, MD2, Fig. 2a, c) that are flanked by rhomboid scales. The first median dorsal plate, as large as the median extrascapular, is about one-third of the second median dorsal plate in length. In internal view, each of these three unpaired bones bears a keel structure as in the median dorsal plate of placoderms<sup>2,40</sup>. In external view, the second median dorsal plate bears a posterodorsally directed spine, like the median dorsal spine of some antiarch placoderms<sup>40</sup>. The median dorsal plates resemble the median ridge scales or fulcral scales preceding the bases of median fins in basal actinopterygians<sup>32,33,41</sup>. However, they differ in that the ridge scales are much smaller and lack strong keel structure. Although the scutes of sturgeons are like the ridge scales, they are obviously derived for actinopterygians<sup>41,42</sup>.

The dorsal fin spine is exceptional in its association with other parts of the body (Fig. 2a, b, d), and forms the leading edge of the dorsal fin. It consists of a stout base with vermiform ridges and a slender,

recurved spine ornamented with about eight longitudinal ridges. Its position in the rear half of the body indicates that the unpaired fin spine of *Psarolepis*<sup>4,17</sup> might be more posteriorly located than previously restored. No fin spine is found for the pelvic and anal fins. A large median ridge scale precedes the dorsal fin spine. Along the ventral surface of the body, a similar large ridge scale is found at a position anterior to the level of the pelvic fins. Between this large scale and the base of the pectoral fin, there are many oval ridge scales. The anal fin is not far from the pelvic fins, and between them are a few paired, elongate, oval cloacal scales (Fig. 2b) as present in *Moythomasia*<sup>32</sup>. The presence of a narrow ventral ridge scale at the posterior end of the holotype is similar to those found near the lower hypochordal lobe of basal actinopterygians and indicates the position of the missing caudal fin. The lepidotrichia are partially preserved in the holotype (Fig. 2a). The large rhomboid flank scales are arranged into at least 30 vertically oriented rows from the median extrascapular to the base of the dorsal fin spine. They bear a prominent anterodorsal process, and are ornamented with linear ridges that terminate in up to 30 posterior serrations. Internally, the scales show the peg-and-socket articulation and dual ridges as in *Ligulalepis*<sup>35</sup>. The isolated lateral-line scales were found in the collection (Fig. 4g–i); however, the path of the lateral line is faintly outlined in the flank of the holotype.

#### Phylogenetic relationships

To determine the placement of *Guiyu* and its impact on osteichthyan relationships, we constructed a new data matrix of 153 characters and

23 taxa including three non-osteichthyan taxa (*Acanthodes*, an acanthodian, *Akmonistion*, a chondrichthyan, and *Dicksonosteus*, a placoderm), mainly revised from references<sup>9, 10, 29, 43</sup>. Phylogenetic analysis (Supplementary Information 1) yields a single most parsimonious tree of length 292, in which *Guiyu* is placed as the sister taxon to the clade comprising *Psarolepis* plus *Achoania*. *Meemannia* and *Ligulalepis* form successively more basal taxa among the Sarcopterygii. *Lophosteus* and *Andreolepis* represent the stem osteichthyans, and *Dialipina* forms the most basal taxon among the Actinopterygii.

### Dating the actinopterygian–sarcopterygian split

The fossil record provides a key means of molecular clock calibration<sup>44,45</sup>. A minimum constraint on the divergence of actinopterygians and sarcopterygians was previously based on the lowest occurrence of crown osteichthyan *Psarolepis* from the top of the Yulungssu Formation<sup>4,17,45</sup> (Fig. 1), dating close to the Silurian/Devonian boundary, ~416 Myr ago. The discovery of *Guiyu*, with accurate dating based on Silurian conodont zonation, provides a new minimum date for molecular-clock-based estimates of the osteichthyan crown node (Fig. 5). As the earliest known, yet not most basal, member of the crown osteichthyan lineage, *Guiyu* lived ~419 Myr ago; thus, the actinopterygian–sarcopterygian split must have occurred no later than this time.

### The mosaic of gnathostome characters in *Guiyu*

Gnathostomes have their possible earliest occurrence in the Late Ordovician<sup>2,46</sup>. The Silurian has yielded definite but fragmentary gnathostome remains, with the only articulated representative being a possible acanthodian from Australia<sup>47</sup> that merely preserves the middle part of the body and adds little novel data to early gnathostome morphology. Until now, the oldest known near-complete gnathostomes were represented by diverse acanthodians<sup>2</sup> and an antiarch placoderm<sup>48</sup> from the earliest Devonian (Lochkovian). The discovery of *Guiyu* provides not only the exceptional corroboration of weak inferences based on disarticulated remains attributed to *Psarolepis* or *Lophosteus*, but also the near-complete restoration of a

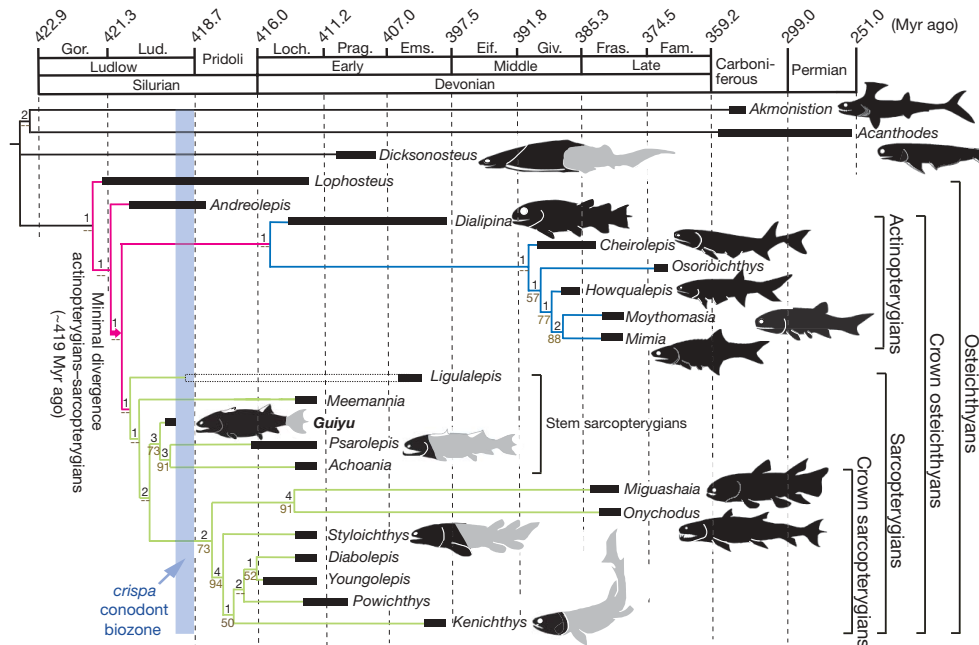
primitive fish with mosaic gnathostome characters. The trunk shield of early osteichthyans, including *Andreolepis*, *Ligulalepis* and *Meemannia*, might bear median fin spine and spine-bearing pectoral girdle as well. Isolated remains that were found in association with these early osteichthyans and referred to non-osteichthyan groups should be reviewed in the light of *Guiyu*.

The phylogenetic position of *Guiyu* and related taxa has a substantial impact on the understanding of key osteichthyan apomorphies. Our phylogenetic scenario, although it should be approached with some caution considering the lower Bremer indices at several nodes, offers new insights into the incremental acquisition of osteichthyan apomorphies, and suggests considerable parallelism between actinopterygians and sarcopterygians. For instance, the placement of *Ligulalepis* at the base of the Sarcopterygii conforms to the notion that its neurocranium reveals the primitive osteichthyan condition<sup>6,7</sup>. *Guiyu*, as well as *Psarolepis* and *Achoania*, is more derived with regard to its dermal and endoskeletal intracranial joints, although it shares many primitive features with non-osteichthyan gnathostomes. The resemblance of the cheek and operculo-gular bone pattern between actinopterygians and stem sarcopterygians denotes a primitive condition for osteichthyans. The anterodorsal process of the scale might be a synapomorphy of crown osteichthyans rather than the actinopterygians.

The discovery of *Guiyu* offers an exceptional example of a primitive fish close to the split of crown osteichthyans. However, our understanding of the stem section of the Osteichthyes phylogenetic tree still remains vague owing to the rarity of relevant fossils. More evidence on Silurian placoderms, acanthodians, chondrichthyans and the fragmentary osteichthyan taxa that are basal to *Guiyu* will provide tests of the present phylogeny and document further the origin of the Osteichthyes.

### METHODS SUMMARY

We adopted the traditional apomorphy-based definition of the Osteichthyes<sup>2</sup>, considering the unresolved deep phylogeny of gnathostomes. Phylogenetic analysis was performed in PAUP4.0b10 (ref. 49) using a branch-and-bound search with default setting, with *Dicksonosteus*, *Akmonistion* and *Acanthodes* specified as the outgroup. Life reconstruction (Supplementary Information 3) was drawn by B. Choo (Victoria Museum) through communication with M.Z.



**Figure 5 | Timing of earliest evolution of the Osteichthyes.** The topology is the most parsimonious tree arising from a matrix of 23 taxa coded for 153 morphological characters (tree length = 292, consistency index = 0.572, retention index = 0.737, rescaled consistency index = 0.421). The numbers at nodes indicate bootstrap support (where the value is greater than 50%)

and Bremer decay index (bottom and top numbers, respectively). The absolute age assessment is from ref. 50. Eif., Eifelian; Ems., Emsian; Fam., Famennian; Fras., Frasnian; Giv., Givetian; Gor., Gorstian; Loch., Lochkovian; Lud., Ludfordian; Prag., Pragian.



Received 5 October 2008; accepted 23 January 2009.

- Nelson, J. S. *Fishes of the World* (Wiley, 2006).
- Janvier, P. *Early Vertebrates* (Clarendon Press, 1996).
- Yu, X.-B. A new porolepiform-like fish *Psarolepis romeri* gen. et sp. nov. (Sarcopterygii, Osteichthyes) from the Lower Devonian of Yunnan, China. *J. Vert. Paleontol.* **18**, 261–274 (1998).
- Zhu, M., Yu, X.-B. & Janvier, P. A primitive fossil fish sheds light on the origin of bony fishes. *Nature* **397**, 607–610 (1999).
- Zhu, M. & Schultze, H.-P. in *Major Events in Early Vertebrate Evolution* (ed. Ahlberg, P. E.) 289–314 (Taylor & Francis, 2001).
- Basden, A. M., Young, G. C., Coates, M. I. & Ritchie, A. The most primitive osteichthyan braincase? *Nature* **403**, 185–188 (2000).
- Basden, A. M. & Young, G. C. A primitive actinopterygian neurocranium from the Early Devonian of southeastern Australia. *J. Vert. Paleontol.* **21**, 754–766 (2001).
- Schultze, H.-P. & Cumbaa, S. L. in *Major Events in Early Vertebrate Evolution* (ed. Ahlberg, P. E.) 315–332 (Taylor & Francis, 2001).
- Zhu, M., Yu, X.-B., Wang, W., Zhao, W.-J. & Jia, L.-T. A primitive fish provides key characters bearing on deep osteichthyan phylogeny. *Nature* **441**, 77–80 (2006).
- Friedman, M. *Styloichthys* as the oldest coelacanth: implications for early osteichthyan interrelationships. *J. Syst. Palaeontol.* **5**, 289–343 (2007).
- Botella, H., Blom, H., Dorka, M., Ahlberg, P. E. & Janvier, P. Jaws and teeth of the earliest bony fishes. *Nature* **448**, 583–586 (2007).
- Gross, W. Fragliche Actinopterygier-Schuppen aus dem Silur Gotlands. *Lethaia* **1**, 184–218 (1968).
- Gross, W. *Lophosteus superbus* Pander, ein Teleostome aus dem Silur Oesels. *Lethaia* **2**, 15–47 (1969).
- Janvier, P. On the oldest known teleostome fish *Andreolepis heder* Gross (Ludlow of Gotland), and the systematic position of the lophosteids. *Eesti NSV Teaduste Akad. Toimetised Geol.* **27**, 88–95 (1978).
- Gagnier, P. Y., Jahnke, H. & Shi, Y. A fish fauna of the Lower Yulongsi Formation (Upper Silurian) of Qujing (E. Yunnan, S. W. China) and its depositional environment. *Cour. Forsch.-Inst. Senckenberg* **110**, 123–135 (1989).
- Wang, N.-Z. & Dong, Z.-Z. Discovery of Late Silurian microfossils of Agnatha and fishes from Yunnan, China. *Acta Palaeont. Sin.* **28**, 192–206 (1989).
- Zhu, M. & Schultze, H.-P. The oldest sarcopterygian fish. *Lethaia* **30**, 293–304 (1997).
- Cumbaa, S. L. & Schultze, H.-P. An Early Devonian (Emsian) acanthodian from the Bear Rock Formation, Anderson River, Northwest Territories, Canada. *Can. J. Earth Sci.* **39**, 1457–1465 (2002).
- Schultze, H.-P. in *Fossil Fishes as Living Animals* (ed. Mark-Kurik, E.) 233–242 (Academy of Sciences of Estonia, 1992).
- Jarvik, E. Middle and Upper Devonian Porolepiformes from East Greenland with special reference to *Glyptolepis groenlandica* n. sp. *Meddr. Grønland* **187**, 1–307 (1972).
- Denison, R. H. Early Devonian lungfishes from Wyoming, Utah, and Idaho. *Fieldiana Geol.* **17**, 353–413 (1968).
- Janvier, P. & Phuong, T. H. Les vertébrés (Placodermi, Galeaspida) du Dévonien inférieur de la coupe de Lung Cô-Mia Lé, province de Hà Giang, Viêt Nam, avec des données complémentaires sur les gisements à vertébrés du Dévonien du Bac Bo oriental. *Geodiversitas* **21**, 33–67 (1999).
- Wang, C.-Y. Age of the Guandi Formation in Qujing District, E. Yunnan. *J. Stratigr.* **25**, 125–127 (2001).
- Ting, V.-K. & Wang, Y.-L. Cambrian and Silurian Formations of Malung and Chutsing Districts, Yunnan. *Bull. Geol. Soc. China* **16**, 1–28 (1937).
- Fang, R.-S. et al. *The Middle Silurian and Early Devonian Stratigraphy and Palaeontology in Qujing District, Yunnan* (Yunnan People's Publishing House, 1985).
- Rong, J.-Y. et al. Some problems concerning the correlation of the Silurian rocks in South China. *J. Stratigr.* **14**, 161–177 (1990).
- Walliser, O. H. & Wang, C.-Y. Upper Silurian stratigraphy and conodonts from the Qujing District, East Yunnan, China. *Cour. Forsch.-Inst. Senckenberg* **110**, 111–121 (1989).
- Pan, J. Notes on Silurian vertebrates of China. *Bull. Chinese Acad. Geol. Sci.* **15**, 227–249 (1986).
- Zhu, M., Yu, X.-B. & Ahlberg, P. E. A primitive sarcopterygian fish with an eyestalk. *Nature* **410**, 81–84 (2001).
- Zhu, M. & Yu, X.-B. in *Recent Advances in the Origin and Early Radiation of Vertebrates* (eds Arratia, G., Wilson, M. V. H. & Cloutier, R.) 271–286 (Verlag Dr. Friedrich Pfeil, 2004).
- Pearson, D. M. & Westoll, T. S. The Devonian actinopterygian *Cheirolepis* Agassiz. *Trans. R. Soc. Edinb.* **70**, 337–399 (1979).
- Gardiner, B. G. The relationships of the palaeoniscid fishes, a review based on new specimens of *Mimia* and *Moythomasia* from the Upper Devonian of western Australia. *Bull. Br. Mus. Nat. Hist.* **37**, 173–428 (1984).
- Long, J. A. New palaeoniscoid fishes from the Late Devonian and Early Carboniferous of Victoria. *Mem. Ass. Australas. Palaeontol.* **7**, 1–64 (1988).
- Friedman, M. & Blom, H. A new actinopterygian from the Famennian of East Greenland and the interrelationships of Devonian ray-finned fishes. *J. Paleontol.* **80**, 1186–1204 (2006).
- Schultze, H.-P. Palaeoniscoiden-Schuppen aus dem Unterdevon Australiens und Kanadas und aus dem Mitteldevon Spitzbergens. *Bull. Br. Mus. Nat. Hist.* **16**, 341–368 (1968).
- Richter, M. & Smith, M. M. A microstructural study of the ganoine tissue of selected lower vertebrates. *Zool. J. Linn. Soc.* **114**, 173–212 (1995).
- Andrews, S. M., Long, J., Ahlberg, P. E., Barwick, R. & Campbell, K. S. W. The structure of the sarcopterygian *Onychodus jandemarra* n. sp. from Gogo, Western Australia: with a functional interpretation of the skeleton. *Trans. R. Soc. Edinb. Earth Sci.* **96**, 197–307 (2006).
- Jarvik, E. *Basic Structure and Evolution of Vertebrates* Vol. 1 (Academic Press, 1980).
- Miller, R. F., Cloutier, R. & Turner, S. The oldest articulated chondrichthyan from the Early Devonian Period. *Nature* **425**, 501–504 (2003).
- Denison, R. H. *Placodermi. Handbook of Paleichthyology* Vol. 2 (ed. Schultze, H.-P.) 1–128 (Gustav Fischer Verlag, 1978).
- Grande, L. & Bemis, W. E. Osteology and phylogenetic relationships of fossil and recent paddlefishes (Polyodontidae) with comments on the interrelationships of Acipenseriformes. *J. Vert. Paleontol.* **11** (Suppl. 1), 1–121 (1991).
- Bemis, W. E., Findeis, E. K. & Grande, L. An overview of Acipenseriformes. *Environ. Biol. Fishes* **48**, 25–71 (1997).
- Zhu, M. & Yu, X.-B. A primitive fish close to the common ancestor of tetrapods and lungfish. *Nature* **418**, 767–770 (2002).
- Müller, J. & Reisz, R. R. Four well-constrained calibration points from the vertebrate fossil record for molecular clock estimates. *Bioessays* **27**, 1069–1075 (2005).
- Benton, M. J. & Donoghue, P. C. J. Paleontological evidence to date the tree of life. *Mol. Biol. Evol.* **24**, 26–53 (2007).
- Sansom, I. J., Smith, M. M. & Smith, M. P. in *Major Events in Early Vertebrate Evolution* (ed. Ahlberg, P. E.) 156–171 (Taylor & Francis, 2001).
- Burrow, C. J. & Young, G. C. An articulated teleostome fish from the Late Silurian (Ludlow) of Victoria, Australia. *Rec. West. Austr. Mus.* **57** (Suppl.), 1–14 (1999).
- Zhang, G.-R., Wang, J.-Q. & Wang, N.-Z. The structure of pectoral fin and tail of Yunnanolepidoidei, with a discussion of the pectoral fin of chuchinolepids. *Vertebr. Palasiat.* **39**, 1–13 (2001).
- Swofford, D. L. *PAUP\*: Phylogenetic Analysis Using Parsimony (\* and other methods), Version 4.0b 10* (Sinauer Associates, 2003).
- Gradstein, F. M. et al. *A Geologic Time Scale 2004* (Cambridge Univ. Press, 2004).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank M.-M. Chang, P. Janvier, J. Long, P. Ahlberg, X.-B. Yu and B. Choo for discussions, X.-B. Yu and B. Choo for improving the manuscript, J. Zhang and C.-H. Xiong for field work, X.-F. Lu and C.-H. Xiong for specimen preparation, W.-D. Zhang and W. Wang for making thin sections, B. Choo for life restoration, J.-L. Huang for illustrations, and J. Zhang for photography. This work was supported by the Major State Basic Research Projects (2006CB806400) of MST of China, the Chinese Academy of Sciences, the National Natural Science Foundation of China, and the CAS/SAFEA International Partnership Program for Creative Research Teams. M.Z., J.L. and T.Q. thank G. Young and J. Long for the examination of Australian specimens with the support of an Australian Research Council Discovery Grant (DP0772138).

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to M.Z. ([zhumin@ivpp.ac.cn](mailto:zhumin@ivpp.ac.cn)).

# Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution

Daniel A. Pomeranz Krummel<sup>1†\*</sup>, Chris Oubridge<sup>1\*</sup>, Adelaine K. W. Leung<sup>1†</sup>, Jade Li<sup>1</sup> & Kiyoshi Nagai<sup>1</sup>

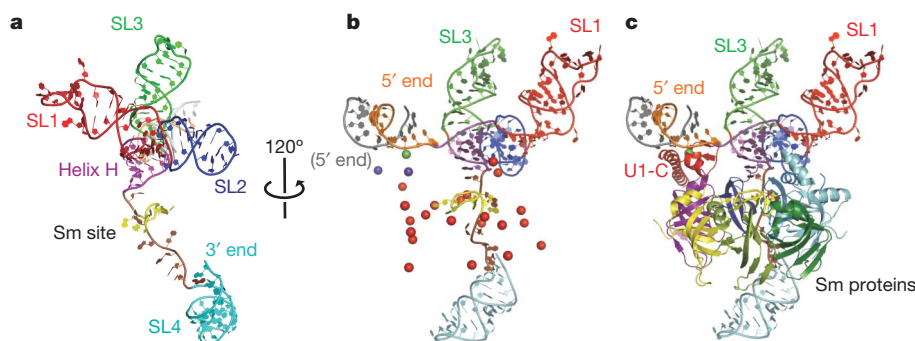
Human spliceosomal U1 small nuclear ribonucleoprotein particles (snRNPs), which consist of U1 small nuclear RNA and ten proteins, recognize the 5' splice site within precursor messenger RNAs and initiate the assembly of the spliceosome for intron excision. An electron density map of the functional core of U1 snRNP at 5.5 Å resolution has enabled us to build the RNA and, in conjunction with site-specific labelling of individual proteins, to place the seven Sm proteins, U1-C and U1-70K into the map. Here we present the detailed structure of a spliceosomal snRNP, revealing a hierarchical network of intricate interactions between subunits. A striking feature is the amino (N)-terminal polypeptide of U1-70K, which extends over a distance of 180 Å from its RNA binding domain, wraps around the core domain consisting of the seven Sm proteins and finally contacts U1-C, which is crucial for 5'-splice-site recognition. The structure of U1 snRNP provides insights into U1 snRNP assembly and suggests a possible mechanism of 5'-splice-site recognition.

In eukaryotes, the majority of protein-coding genes are interrupted by non-coding sequences known as introns. The entire length of a gene, including its introns, is transcribed into precursor mRNAs (pre-mRNAs). The introns are excised and the exons spliced together to form mRNA with continuous protein-coding sequences by a large RNA–protein assembly called the spliceosome<sup>1–3</sup>. The major components of the spliceosome are five types of snRNP (U1, U2, U4, U5 and U6), each containing one of five spliceosomal U-type small nuclear RNAs (snRNAs), seven Sm or Lsm proteins and particle-specific proteins<sup>1–6</sup>. Assembly of the spliceosome is initiated by the binding of U1 snRNP to the 5' splice site in pre-mRNA, followed by an ATP-dependent binding of U2 snRNP to the branch-point sequence within the intron. Upon further binding of the tri-snRNP, which contains U4, U6 and U5 snRNPs, the spliceosome undergoes extensive rearrangements to become catalytically active<sup>1–3,5,6</sup>.

Mammalian U1 snRNP consists of U1 snRNA, seven Sm proteins (Sm-B/Sm-B', Sm-D1, Sm-D2, Sm-D3, Sm-E, Sm-F and Sm-G) and

three U1-specific proteins (U1-70K, U1-A and U1-C). The base-pairing between the 5' end of U1 snRNA<sup>7,8</sup> and the 5' splice site in pre-mRNA has a crucial role in 5'-splice-site recognition<sup>9–11</sup>. In higher eukaryotes, splicing factors, such as SR proteins, bound to splice enhancer sequences direct U1 snRNP through protein–protein interactions<sup>12,13</sup> to the vicinity of 5' splice sites that differ significantly from the consensus sequence<sup>1</sup>.

Human U1 snRNA forms four stem-loops<sup>14</sup> (Supplementary Fig. 1). Seven Sm proteins assemble around the Sm site nucleotides, located between stem-loop 3 (SL3) and SL4, to form the core domain<sup>3,4</sup>. U1-70K and U1-A bind to SL1 and SL2, respectively, through their RNA binding domains (RBDs)<sup>15–19</sup>. U1-C does not bind free U1 snRNA, but the addition of the first 97 residues of U1-70K enables U1-C to bind to the U1 core domain<sup>20</sup>. We have previously reported the crystal structure of the N-terminal RBD of U1-A bound to SL2 of U1 snRNA<sup>19</sup> and the solution structure of the zinc-finger domain of U1-C<sup>21</sup>. We proposed a heptameric ring model of the core domain based on the crystal structures of two Sm protein heterodimers (Sm-D3–Sm-B and



**Figure 1 | Overall fold of U1 snRNA and the U1 snRNP core domain containing the seven Sm proteins.** **a**, Human U1 snRNA forms four stem-loop structures: SL1 (red), SL2 (blue), SL3 (green) and SL4 (cyan). A four-helix junction with two coaxially stacked helices (SL1–SL2 and SL3–helix H (magenta)) is separated from SL4 by the single-stranded Sm site (yellow-green). **b**, Anomalous difference peaks used to place the seven Sm proteins

and U1-C into the map. Selenium atoms of seleno-methionine (SeMet)-labelled Sm proteins (red); a zinc atom of the zinc-finger of U1-C (green); mercury atoms attached to residues Cys 24 and mutant Q39C of U1-C (purple). The 5' end of U1 snRNA (orange) base-pairs with its counterpart from a neighbouring complex (grey). **c**, Seven Sm proteins and U1-C fitted into the map.

<sup>1</sup>MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK. <sup>†</sup>Present addresses: Department of Biochemistry, Brandeis University, Waltham, Massachusetts 02454-9110, USA (D.A.P.K.); Department of Neurobiology, Harvard Medical School, Boston, Massachusetts 02115, USA (A.K.W.L.).

\*These authors contributed equally to this work.



Sm-D1–Sm-D2)<sup>22</sup>. However, the three-dimensional organization of these components in U1 snRNP is not known.

Early electron microscopy studies of negatively stained U1 snRNP revealed a globular core domain with two protuberances attributable to U1-70K and U1-A<sup>23,24</sup>. More recently, Stark *et al.*<sup>25</sup> carried out a cryo-electron microscopy single-particle analysis of U1 snRNP, but the resolution was not sufficient to resolve secondary structure elements of proteins. Therefore, the authors were unable to place protein components unambiguously and the proposed arrangement of proteins thus relied largely on biochemical data.

We have crystallized and solved the structure of the functional core of human U1 snRNP at 5.5 Å resolution. At this resolution, the major and minor grooves of double-stranded RNA are evident. In addition,  $\alpha$ -helices of proteins can be seen directly and  $\beta$ -sheets appear as flat density<sup>26,27</sup>, allowing known protein folds to be identified. We have

specifically labelled residues in individual proteins and used anomalous difference peaks from selenium, zinc and mercury to place each protein unambiguously into the map. This also allowed us to trace the N-terminal 100 residues of U1-70K with no prior structural information. The crystal structure of human U1 snRNP presented here reveals the detailed three-dimensional organization of the RNA and protein components within a spliceosomal snRNP and provides important insight into its function.

### Structure determination of U1 snRNP

We reconstituted U1 snRNP from proteins expressed in *Escherichia coli* and U1 snRNA produced by *in vitro* transcription<sup>19,21,22</sup>. U1 snRNP is active in a splicing assay when U1-A is depleted<sup>28</sup> or nucleotides 50–91 of U1 snRNA, which contain the U1-A binding site, are deleted<sup>29</sup>. Crystals of the functional core of U1 snRNP were obtained with Sm-B (1–174), U1-C (1–77), U1-70K (2–215), the full-length constructs of Sm-D1, Sm-D2, Sm-D3, Sm-E, Sm-F and Sm-G, and U1 snRNA in which the apical part of SL2 was replaced with a kissing-loop sequence<sup>30</sup> to promote crystal contacts (Supplementary Fig. 1). The crystals belong to the triclinic *P*1 space group with typical unit-cell dimensions of  $a = 127$  Å,  $b = 128$  Å,  $c = 156$  Å,  $\alpha = 96^\circ$ ,  $\beta = 107^\circ$  and  $\gamma = 101^\circ$  and four complexes in the asymmetric unit. Initial phases were obtained with a Ta<sub>6</sub>Br<sub>12</sub> derivative. After inclusion of other heavy-atom derivatives, an electron density map at a resolution of 6.5 Å allowed the building of all RNA helices and fitting of known protein domains. A mercury-soaked crystal containing a U1-C variant (Q39C) diffracted to 5.5 Å. After multi-domain, multi-crystal averaging<sup>31</sup>, the map calculated with the 5.5 Å data set was of excellent quality (Supplementary Table 1).

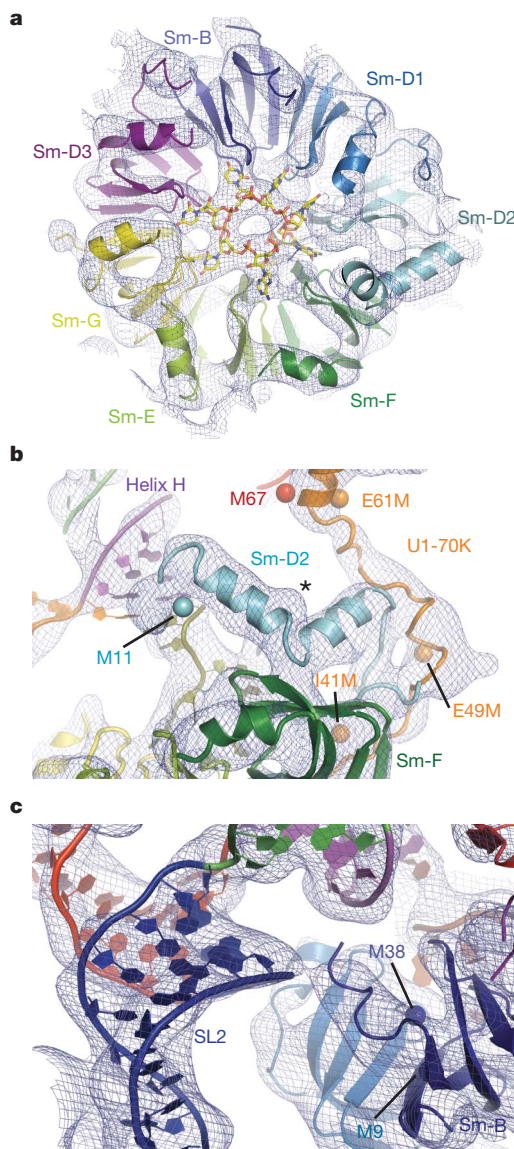
### Structure of the U1 snRNP core domain

RNA on the 5' side of the Sm site forms a four-helix junction and SL4 is located on the opposite side of the Sm ring (Fig. 1). SL1 and SL2 coaxially stack on each other. Similarly, SL3 and helix H stack coaxially (Fig. 1a). These coaxially stacked helices cross at an angle of  $\sim 90^\circ$ . This stacking scheme was proposed for free U1 snRNA on the basis of biochemical studies<sup>14,32</sup>.

The Sm ring consisting of seven Sm proteins can be recognized in the electron density where selenium anomalous peaks from all Sm proteins are clustered (Fig. 1b). The identity of the seven Sm proteins was established unambiguously from the positions of these anomalous peaks. The Sm proteins are arranged in the order Sm-E, Sm-D3, Sm-B, Sm-D1, Sm-D2, Sm-F (Figs 1c and 2a), in agreement with our heptameric ring model<sup>22</sup>. The averaged map of 5.5 Å resolution is of sufficient quality that a rod-like density for the N-terminal helix and a curved slab of density for the  $\beta$ -sheet of the Sm fold are almost perfectly formed in each Sm protein (Fig. 2a).

Sm-D2 contains a particularly long N-terminal extension, whereas Sm-B, Sm-D1 and Sm-D3 have long carboxy (C)-terminal extensions. The four-helix junction of U1 snRNA lies over a relatively flat face of the Sm protein ring consisting of the N-terminal helices of the Sm proteins. The N terminus of D2 is not ordered in the Sm-D1–Sm-D2 heterodimer<sup>22</sup>, but in U1 snRNP it forms a kinked helix and interacts with helix H in its minor groove (Fig. 2b). The positive end of the  $\alpha$ -helix dipole, as well as the conserved Lys 6 and Lys 8 residues, may facilitate the interaction of the N terminus of Sm-D2 with RNA. The N terminus of Sm-B also interacts with the RNA backbone at the base of SL2 (Fig. 2c). These two interactions stabilize the four-helix junction of U1 snRNA onto the core domain and stabilize the overall structure of U1 snRNP.

A cartwheel-shaped electron density observed in the central hole of the Sm ring (Fig. 2a) is attributed to the Sm site RNA in analogy to the density observed for the pentaurydylate bound to the homoheptamer of Lsm proteins<sup>33</sup>. This density has features of bases splayed out towards the Sm proteins. The first and third uracil bases within an Sm-site nonamer oligonucleotide (5'-rAAUUUUUGG-3') were crosslinked to Sm-G and Sm-B, respectively<sup>34</sup>. These two positions correspond to



**Figure 2 | Structure of the U1 snRNP core domain.** **a**, The Sm proteins and the experimental electron density map (contoured at  $1\sigma$ ). Cartwheel-shaped density in the central hole of the U1 snRNP core domain is attributed to the seven Sm site nucleotides. **b**, The N terminus of Sm-D2 supporting helix H. The extended polypeptide of U1-70K interacts with Sm-D2. SeMet anomalous peaks from natural (Met 67; red) and engineered Met residues (I41M, E49M and E61M; orange) of U1-70K and of Sm-D2 (Met 11; cyan). The asterisk indicates a 'bump' due to Phe 24 of Sm-D2. **c**, The N terminus of Sm-B interacting with the phosphate backbone of SL2. Navy blue spheres, anomalous peaks from SeMet-labelled Sm-B.

U127 and U129 of U1 snRNA and the register of RNA binding within the Sm ring can be inferred on the basis of these two crosslinks. Thus, the seven nucleotides of the Sm site (5'-AUUUGUG-3') are likely to interact with Sm-E, Sm-G, Sm-D3, Sm-B, Sm-D1, Sm-D2 and Sm-F, respectively.

### The zinc-finger domain of U1-C

The zinc-finger domain of U1-C<sup>21</sup> interacts with the Sm protein ring exclusively through Sm-D3 (Fig. 1c). In solution helix C of U1-C folds back onto helix B through clustering of aromatic amino acids<sup>21</sup>, whereas in U1 snRNP these helices form a long continuous helix (helix B) extending from the Zn-binding site (Fig. 3a). In the crystal the 5' end of U1 snRNA base-pairs with its counterpart from an adjacent complex (Fig. 1b). The proposed base-pairing scheme was inferred from the length of the RNA duplex (Fig. 3c). This interaction is likely to mimic the binding of the 5' splice site of pre-mRNA to U1 snRNA (Fig. 3d). Helix A of U1-C binds across the minor groove of this RNA duplex where C8 and A7 of U1 snRNA base-pair with nucleotides corresponding to the invariant GU dinucleotide of the 5' splice site (Fig. 3a). Importantly, the zinc-finger of U1-C is distinct from the TFIIIA-type<sup>35</sup> in that there are five intervening residues between the two zinc-coordinating histidines. To allow these two histidines to coordinate the zinc ion, the intervening residues in U1-C form a loop. This atypical zinc-finger structure<sup>21</sup> places the loop and helix A at strategic positions to interact with the duplex between the 5' end of U1 snRNA and the putative pre-mRNA strand (Fig. 3a).

### U1-70K has an unusual structure and path

U1-70K contains an RBD between residues 100 and 180, whereas its N-terminal region (residues 1–100) has no known sequence motif. We placed a homology-modelled RBD of U1-70K into a large density near SL1 of U1 snRNA with the aid of two selenium positions (Met 134 and Met 157; Fig. 4a). On the basis of the crosslinking data<sup>36</sup>, we built the RNA loop of SL1 into the density in such a way that G28 and U30 of SL1 were close to Tyr 112 and Leu 175, respectively. The region between residues 61 and 89 of U1-70K, predicted to form an  $\alpha$ -helix, was built into the long tubular density extending along the RNA stem from the RBD (Fig. 4a). The register and orientation of the  $\alpha$ -helix were determined from the selenium peaks of Met 67 and Met 88. The region around the Met 67 and I75M landmarks contains nine basic residues. Of these, Arg 63, Arg 66, Arg 70, Arg 71, Lys 74, Arg 68 and Arg 79 are

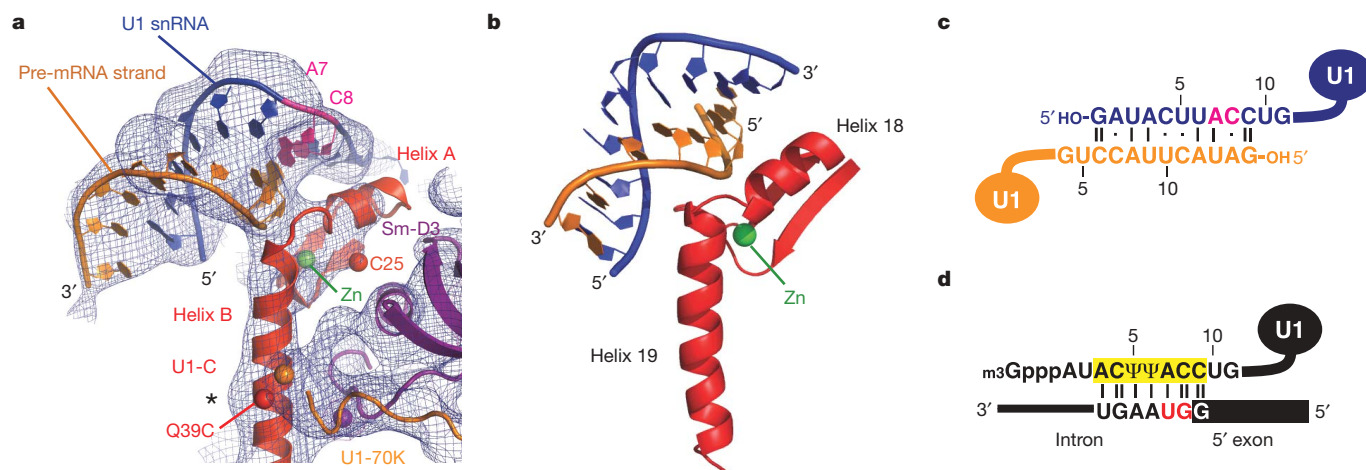
close to the phosphate backbone of SL1. Hence, these residues are likely to have an important role in positioning this long  $\alpha$ -helix along SL1.

The N-terminal 97 residues of U1-70K are necessary and sufficient for U1-C to bind to the U1 core domain<sup>20</sup>. The RBD and the long  $\alpha$ -helix of U1-70K, distant from U1-C, do not provide a plausible explanation for the essential role of U1-70K in U1-C binding. To trace the path of the N terminus, crystals of seven variants of U1 snRNP were grown, each with a SeMet derivative of single methionine mutations of U1-70K (L9M, I19M, E31M, I41M, E49M, E61M and I75M). These SeMet anomalous peaks unambiguously established the path of the N-terminal region of U1-70K (Fig. 4b–d). Notably, the extended polypeptide chain of U1-70K wraps around the core domain and reaches an  $\alpha$ -helical density near the Se peak of the L9M mutant. Electron density attributable to U1-70K can be identified along this path from the RBD to U1-C (Figs 2b and 4b). The N terminus of U1-70K and the C terminus of Sm-D3 create a binding groove for the long  $\alpha$ -helix of U1-C (Fig. 4b). This interaction accounts for the essential role of the N terminus of U1-70K for U1-C binding<sup>20</sup>.

### The 5'-splice-site recognition

The fortuitous interaction between the 5' ends of U1 snRNA from two U1 snRNPs provides important insights into the mechanism of 5'-splice-site recognition. U1-C is crucial for the formation of E complex<sup>28</sup> in which the 5' splice site of pre-mRNA base-pairs with the 5' end of U1 snRNA<sup>9–11</sup>. The structure of U1 snRNP suggests a role for U1-C in stabilizing this base-pairing. A double mutant of U1-C (R28G, K29S) fails to promote the E-complex formation<sup>29</sup>. These residues are within the loop between helices A and B, making them ideally located to interact with the putative pre-mRNA strand, consistent with their crucial role in 5'-splice-site recognition (Fig. 3a). In the crystal structure of dimethylallyltransferase transfer RNA complex, a zinc-finger domain interacts with the anticodon stem of tRNA<sup>37</sup> (Fig. 3b). This zinc-finger also contains a loop formed by five intervening residues between the two zinc-coordinating histidines, and the loop and the two flanking helices interact with the anticodon stem. The remarkable similarities in RNA binding between the two zinc-fingers support the notion that the zinc-finger of U1-C interacts with and stabilizes the RNA duplex between the 5' end of U1 snRNA and the 5' splice site observed in our crystal structure. The crosslinking between U1-C and the 5' splice site is consistent with the structure<sup>38</sup>.

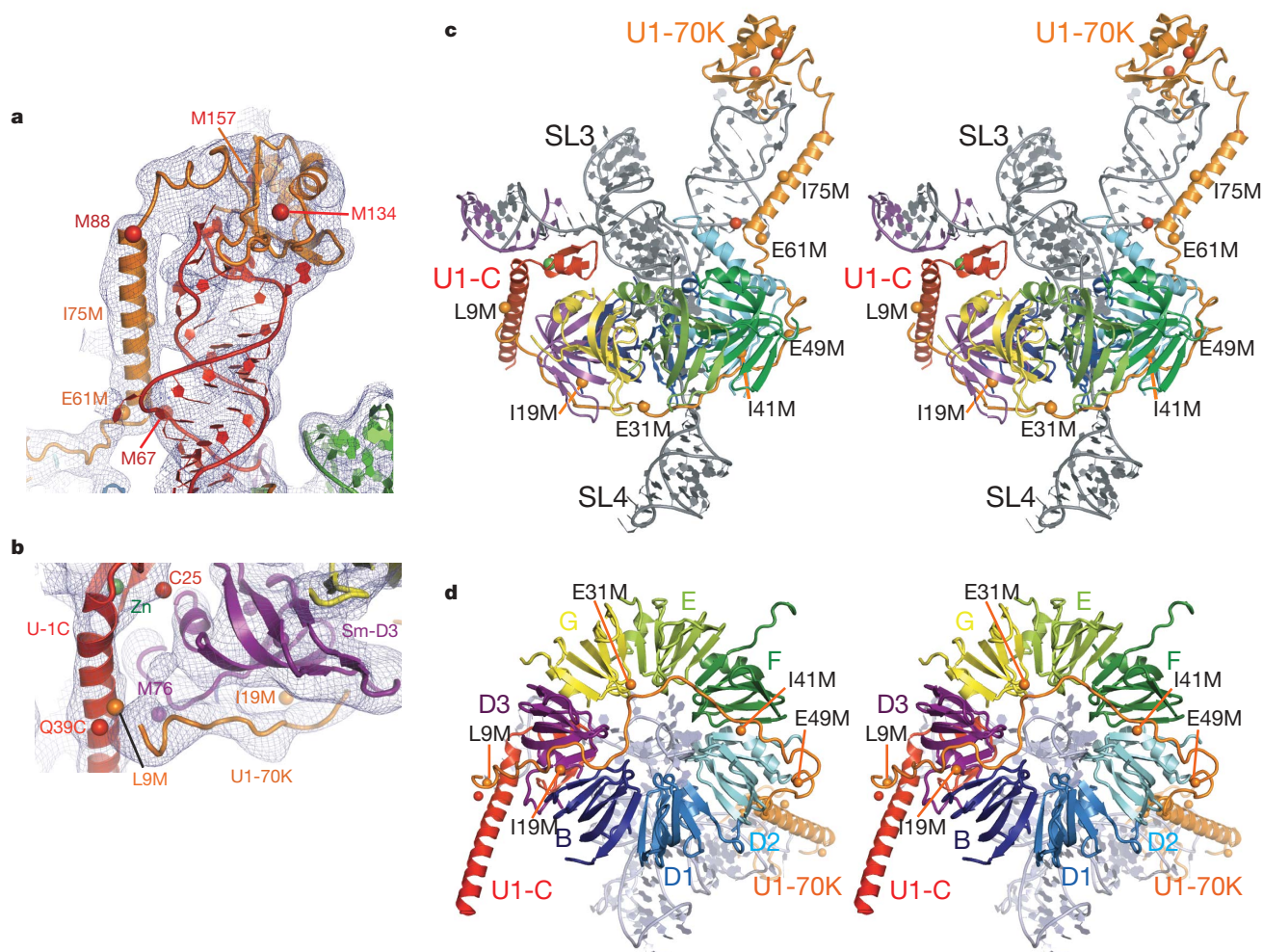
The first and second nucleotides of the intron (the invariant GU dinucleotide) form Watson–Crick base pairs with C8 and A7 of U1



**Figure 3 | A mechanism of 5'-splice-site recognition.** **a**, The 5' end of U1 snRNA (blue) base-pairs with the same region of a neighbouring complex (orange), which mimics the 5' splice site of pre-mRNA. C8 and A7 of U1 snRNA are highlighted (magenta). The experimental map is contoured at 1 $\sigma$ . Spheres represent anomalous peaks of zinc (green) and mercury (red). The asterisk indicates a 'bump' due to Trp 41. **b**, The zinc-finger of yeast

dimethylallyltransferase and the anticodon stem of a tRNA substrate<sup>41</sup>. **c**, Base-pairing between the 5' ends of two U1 snRNAs within the U1 snRNP crystal. C8 and A7 are highlighted (magenta). **d**, Base-pairing between the 5' end of U1 snRNA and a canonical 5' splice site<sup>1</sup>. The invariant first and second nucleotides of the intron (GU) are highlighted (red).  $\Psi$ , pseudouridine.





**Figure 4 | Interaction of U1-70K with SL1 of U1 snRNA, the core domain and U1-C.** **a**, The RBD of U1-70K and the  $\alpha$ -helix extending along SL1 towards the core domain. **b**, The long  $\alpha$ -helix B of U1-C (red) held between the N terminus of U1-70K (orange) and the C-terminal end of Sm-D3 (purple). **c**, Stereo view of U1 snRNP showing the entire path of U1-70K revealed by SeMet landmarks

(orange spheres). **d**, Path of the N-terminal region of U1-70K crossing the Sm ring, viewed in stereo from SL4 (omitted for clarity). The Sm proteins are indicated without the 'Sm-' prefix, for clarity. Spheres indicate anomalous peaks from SeMet in U1-70K (engineered, orange) and Hg attached to U1-C (Q39C) (red). Experimental electron density map is contoured at  $1\sigma$ .

snRNA, respectively (Fig. 3a, d). An intriguing possibility is that the conserved residues in the loop region and helix A of U1-C probe hydrogen-bonding groups of these two base pairs in the minor groove and contribute to a discrimination of nucleotides at these two positions. In yeast, the function of an otherwise essential DEAD-box helicase, Prp28p, can be bypassed by mutations in U1 snRNA or U1-C<sup>39</sup>. It is plausible that the binding of U1-C to a correctly paired 5' splice site and U1 snRNA could be coupled to an activation of a DEAD-box helicase within B complex to allow a transfer of the 5' splice site from U1 snRNA to U6 snRNA<sup>40</sup>. Such a proof-reading mechanism to discriminate incoming nucleotide substrates in the minor groove has been observed for both DNA and RNA polymerases<sup>41</sup>.

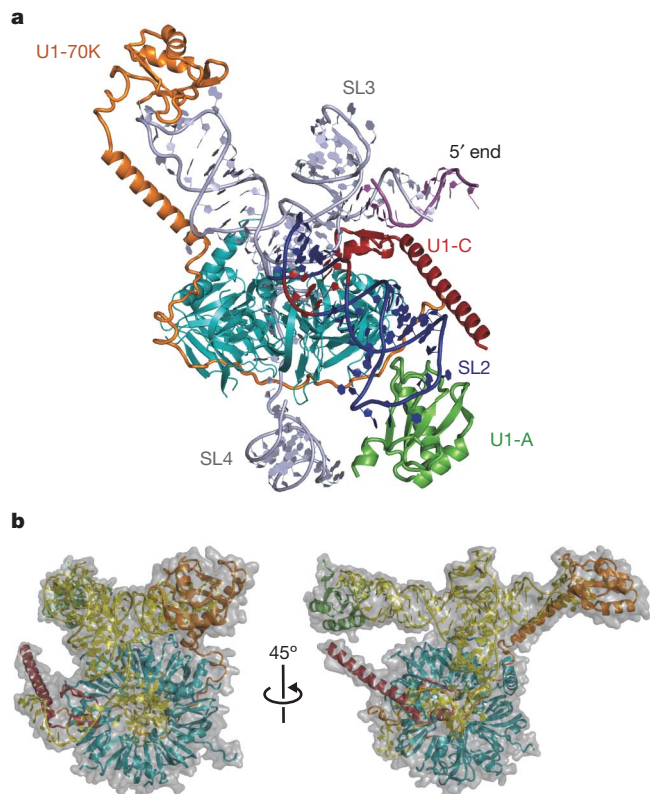
#### Implications for assembly of snRNP

The crystal structure reported here represents the functional core of U1 snRNP, lacking the apical region of SL2. U1-A consists of two RBDs linked by a proline-rich sequence; the N-terminal RBD binds to the ten-nucleotide loop of SL2<sup>18,19</sup>. To complete the model of U1 snRNP, the crystal structure of the U1-A RNA complex<sup>19</sup> was added onto an extended SL2 helix (Fig. 5a). This places the internal loop in SL2 consisting of four conserved non-canonical base pairs (Supplementary Fig. 1) in a position to interact with Sm-B and Sm-D1 on the rim of the Sm ring (Fig. 5a). This interaction could

further stabilize the RNA structure onto the core domain. There is no known function for the C-terminal RBD of U1-A nor for the highly similar C-terminal RBD of U2-B'', a component of U2 snRNP<sup>18,42</sup>. The C-terminal RBDs of these proteins may bind to a common binding partner, possibly the Sm ring.

The cryo-electron microscopy structure of human U1 snRNP<sup>25</sup> shows a ring-shaped core domain with two large protuberances that were previously evident in negatively stained images of U1 snRNP<sup>23,24</sup>. One striking feature in the cryo-electron microscopy structure is the funnel-shaped central hole in the Sm ring. In the crystal structure, a large RNA structure that includes the four-helix junction is located directly above the centre of the Sm ring and no funnel-shaped hole is observed. No class averages were presented in the cryo-electron microscopy paper<sup>25</sup>, but a gallery of negatively stained images of U1 snRNP<sup>23,24</sup> provides different views of U1 snRNP. Closely matching images of these views (fig. 3 in ref. 23) can be generated by rotating our model of the complete U1 snRNP (Fig. 5b).

The crystal structure of a spliceosomal snRNP presented here provides insights into general principles of snRNP assembly. As observed in the U1 snRNP structure, the Sm protein ring assembled on U2, U4 and U5 snRNAs is likely to function as a platform for further protein assembly<sup>1,3</sup>. Functionally important RNA sequences, which make up the catalytic centre of the spliceosome together with



**Figure 5 | Model of the complete human U1 snRNP.** **a**, Overview of a model of the complete U1 snRNP. Truncated SL2 was extended with A-form RNA and the crystal structure of the U1A–RNA complex<sup>19</sup> was appended to the extended helix. The internal loop of SL2, consisting of four consecutive non-Watson–Crick base pairs (red), is in a position to interact with Sm-B and Sm-D1. **b**, Two views of the complete U1 snRNP model approximately 45° apart, with surface representation superimposed. Closely matching images are found in the gallery of negatively stained images of U1 snRNP reported previously<sup>23,24</sup>.

pre-mRNA substrates<sup>1–3</sup>, are located on the 5' side of the Sm or Lsm protein binding site in all spliceosomal snRNAs<sup>1,3</sup>. In U1 snRNP, the region of U1 snRNA that includes the four-helix junction and the 5' end are stabilized in a particular orientation relative to the Sm ring by the N-terminal helices of Sm-D2 and Sm-B (Fig. 2b, c). The N-terminal helix of Sm-D2, not ordered in the Sm-D1–Sm-D2 heterodimer<sup>22</sup>, becomes ordered through its interaction with helix H of U1 snRNA. In U2, U4 and U5 snRNAs, different RNA structures flank the Sm site on the 5' side<sup>1,3</sup>. Hence, the N-terminal helices of Sm-D2 and Sm-B are likely to play a similar role in stabilizing different RNA structures in other spliceosomal snRNPs.

U1-C predominantly interacts with Sm-D3, but the additional interaction with the N terminus of U1-70K is crucial in promoting its binding to the core domain. Assembly of U1 snRNP therefore occurs in a hierarchical process, as in the ribosome<sup>43</sup>. U1-70K facilitates the binding of U1-C by providing an additional interaction surface for U1-C. The N-terminal polypeptide of U1-70K spans a very long distance. One of its functions may be to restrict the movement of SL1 with respect to the core domain, thus stabilizing the overall structure of U1 snRNP. Other snRNP specific proteins may have an analogous role.

The structure presented here reveals the architecture of U1 snRNP, held together by an intricate network of RNA–protein and protein–protein interactions. It has also provided important insights into the mechanisms of 5'-splice-site selection by U1 snRNP. During different stages of spliceosomal assembly, U1 snRNP interacts with other snRNPs either directly or through other proteins<sup>1–3</sup>. Our structure can now be used to study and provide an understanding of the interaction of U1 snRNP with constitutive and alternative splicing factors<sup>12,13,44</sup>, as well as other snRNP<sup>45</sup>, during assembly of the spliceosome.

## METHODS SUMMARY

**Crystallization of U1 snRNP.** The preparation of U1 snRNP proteins and U1 snRNA, and the reconstitution and purification of U1 snRNP, have been described previously<sup>19–21</sup>. The eluate of U1 snRNP from a MonoQ column was concentrated to 5 mg ml<sup>−1</sup>, mixed with an equal volume of a reservoir solution (MES–KOH, pH 6.2–6.6; 300 mM KCl; 38–42% MPD) and equilibrated with the reservoir solution at 4 °C by the hanging-drop method. Crystals suitable for data collection were grown by streak-seeding using a feline whisker at 4 °C. Crystals were collected and cryo-cooled directly from mother liquor.

**Data collection.** Data used for structure determination were collected on the X06SA and X10SA beamlines of the Swiss Light Source (Paul Scherrer Institute) at 100 K on a mar225 charge-coupled-device detector. For all crystals, data sets were obtained with the beam focused on the detector (de-focused) and using inverse beam mode such that Friedel pairs could be collected with similar radiation damage.

**Structural determination.** Data processing, phasing and modelling are as described in Supplementary Table 1.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 10 November 2008; accepted 4 February 2009.

- Burge, C. B., Tuschl, T. & Sharp, P. A. in *The RNA World II* (eds Gesteland, R. F., Cech, T. R. & Atkins, J. F.) 525–560 (Cold Spring Harbor Laboratory Press, 1999).
- Will, C. L. & Lührmann, R. in *The RNA World* (eds Gesteland, R. F., Cech, T. R. & Atkins, J. F.) 369–400 (Cold Spring Harbor Laboratory Press, 2006).
- Yu, Y.-T., Scharl, E. C., Smith, C. M. & Steitz, J. A. in *The RNA World II* (eds Gesteland, R. F., Cech, T. R. & Atkins, J. F.) 487–524 (Cold Spring Harbor Laboratory Press, 1999).
- Bringmann, P. & Lührmann, R. Purification of the individual snRNPs U1, U2, U5 and U4/U6 from HeLa cells and characterization of their protein constituents. *EMBO J.* **5**, 3509–3516 (1986).
- Jurica, M. S. & Moore, M. J. Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell* **12**, 5–14 (2003).
- Makarov, E. M. *et al.* Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science* **298**, 2205–2208 (2002).
- Guthrie, C. & Patterson, B. Spliceosomal snRNAs. *Annu. Rev. Genet.* **22**, 387–419 (1988).
- Griffiths-Jones, S. *et al.* Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* **33**, D121–D124 (2005).
- Zhuang, Y. & Weiner, A. M. A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. *Cell* **46**, 827–835 (1986).
- Ruby, S. W. & Abelson, J. An early hierarchic role of U1 small nuclear ribonucleoprotein in spliceosome assembly. *Science* **242**, 1028–1035 (1988).
- Séraphin, B. & Rosbash, M. Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing. *Cell* **59**, 349–358 (1989).
- Kohtz, J. D. *et al.* Protein–protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature* **368**, 119–124 (1994).
- Graveley, B. R. Sorting out the complexity of SR protein functions. *RNA* **6**, 1197–1211 (2000).
- Krol, A. *et al.* Solution structure of human U1 snRNA. Derivation of a possible three-dimensional model. *Nucleic Acids Res.* **18**, 3803–3811 (1990).
- Patton, J. R. & Pederson, T. The M, 70,000 protein of the U1 small nuclear ribonucleoprotein particle binds to the 5' stem-loop of U1 RNA and interacts with Sm domain proteins. *Proc. Natl Acad. Sci. USA* **85**, 747–751 (1988).
- Query, C. C., Bentley, R. C. & Keene, J. D. A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. *Cell* **57**, 89–101 (1989).
- Bach, M., Krol, A. & Lührmann, R. Structure-probing of U1 snRNPs gradually depleted of the U1-specific proteins A, C and 70k. Evidence that A interacts differentially with developmentally regulated mouse U1 snRNA variants. *Nucleic Acids Res.* **18**, 449–457 (1990).
- Scherly, D., Boelens, W., Dathan, N. A., van Venrooij, W. J. & Mattaj, I. W. Major determinants of the specificity of interaction between small nuclear ribonucleoproteins U1A and U2B' and their cognate RNAs. *Nature* **345**, 502–506 (1990).
- Oubridge, C., Ito, N., Evans, P. R., Teo, C. H. & Nagai, K. Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372**, 432–438 (1994).
- Nelissen, R. L., Will, C. L., van Venrooij, W. J. & Lührmann, R. The association of the U1-specific 70K and C proteins with U1 snRNPs is mediated in part by common U snRNP proteins. *EMBO J.* **13**, 4113–4125 (1994).
- Muto, Y. *et al.* The structure and biochemical properties of the human spliceosomal protein U1C. *J. Mol. Biol.* **341**, 185–198 (2004).
- Kambach, C. *et al.* Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell* **96**, 375–387 (1999).



23. Kastner, B. & Lührmann, R. Electron microscopy of U1 small nuclear ribonucleoprotein particles: shape of the particle and position of the 5' RNA terminus. *EMBO J.* **8**, 277–286 (1989).
24. Kastner, B., Kornstädt, U., Bach, M. & Lührmann, R. Structure of the small nuclear RNP particle U1: identification of the two structural protuberances with RNP-antigens A and 70K. *J. Cell Biol.* **116**, 839–849 (1992).
25. Stark, H., Dube, P., Lührmann, R. & Kastner, B. Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature* **409**, 539–542 (2001).
26. Clemons, W. M. *et al.* Structure of a bacterial 30S ribosomal subunit at 5.5 Å resolution. *Nature* **400**, 833–840 (1999).
27. Ban, N. *et al.* Placement of protein and RNA structures into a 5 Å-resolution map of the 50S ribosomal subunit. *Nature* **400**, 841–847 (1999).
28. Heinrichs, V., Bach, M., Winkelmann, G. & Lührmann, R. U1-specific protein C needed for efficient complex formation of U1 snRNP with a 5' splice site. *Science* **247**, 69–72 (1990).
29. Will, C. L. *et al.* *In vitro* reconstitution of mammalian U1 snRNPs active in splicing: the U1-C protein enhances the formation of early (E) spliceosomal complexes. *Nucleic Acids Res.* **24**, 4614–4623 (1996).
30. Kim, C. H. & Tinoco, I. Jr. A retroviral RNA kissing complex containing only two G-C base pairs. *Proc. Natl Acad. Sci. USA* **97**, 9396–9401 (2000).
31. Cowtan, K. D., Zhang, K. Y. J. & Main, P. in *International Tables for Crystallography* Vol. F: *Crystallography of Biological Macromolecules* (eds Rossmann, M. G. & Arnold, E.) 705–710 (Kluwer, 2001).
32. Duckett, D. R., Murchie, A. I. H. & Lilley, D. M. J. The global folding of four-way helical junctions in RNA, including that in U1 snRNA. *Cell* **83**, 1027–1036 (1995).
33. Törö, I. *et al.* RNA binding in an Sm core domain: X-ray structure and functional analysis of an archaeal Sm protein complex. *EMBO J.* **20**, 2293–2303 (2001).
34. Urlaub, H., Raker, V. A., Kostka, S. & Lührmann, R. Sm protein-Sm site RNA interactions within the inner ring of the spliceosomal snRNP core structure. *EMBO J.* **20**, 187–196 (2001).
35. Lu, D., Searles, M. A. & Klug, A. Crystal structure of a zinc-finger-RNA complex reveals two modes of molecular recognition. *Nature* **426**, 96–100 (2003).
36. Urlaub, H., Hartmuth, K., Kostka, S., Grell, G. & Lührmann, R. A general approach for identification of RNA-protein cross-linking sites within native human spliceosomal small nuclear ribonucleoproteins (snRNPs). *J. Biol. Chem.* **275**, 41458–41468 (2000).
37. Zhou, C. & Huang, R. H. Crystallographic snapshots of eukaryotic dimethylallyltransferase acting on tRNA: Insight into tRNA recognition and reaction mechanism. *Proc. Natl Acad. Sci. USA* **105**, 16142–16147 (2008).
38. Du, H. & Rosbash, M. The U1 snRNP protein U1C recognizes the 5' splice site in the absence of base pairing. *Nature* **419**, 86–90 (2002).
39. Chen, J. Y.-F. *et al.* Specific alterations of U1-C protein or U1 small nuclear RNA can eliminate the requirement of Prp28p, an essential DEAD box splicing factor. *Mol. Cell* **7**, 227–232 (2001).
40. Ismaili, N., Sha, M., Gustafson, E. H. & Konarska, M. M. The 100-kDa U5 snRNP protein (hPrp28p) contacts the 5' splice site through its ATPase site. *RNA* **7**, 182–193 (2001).
41. Steitz, T. A. Visualizing polynucleotide polymerase machines at work. *EMBO J.* **25**, 3458–3468 (2006).
42. Price, S. R., Evans, P. R. & Nagai, K. Crystal structure of the spliceosomal U2B''–U2A' protein complex bound to a fragment of U2 small nuclear RNA. *Nature* **394**, 645–650 (1998).
43. Mizushima, S. & Nomura, M. Assembly mapping of 30S ribosomal proteins from *E. coli*. *Nature* **226**, 1214–1218 (1970).
44. Bedford, M. T., Reed, R. & Leder, P. WW domain-mediated interactions reveal a spliceosome-associated protein that binds a third class of proline-rich motif: The proline, glycine and methionine-rich motif. *Proc. Natl Acad. Sci. USA* **95**, 10602–10607 (1998).
45. Dönmez, G., Hartmuth, K., Kastner, B., Will, C. L. & Lührmann, R. The 5' end of U2 snRNA is in close proximity to U1 and functional sites of the pre-mRNA in early spliceosomal complexes. *Mol. Cell* **25**, 399–411 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We are grateful to T. Jessen, C. Kambach, J. Avis, R. Young, Y. Muto, S. Walke and T. Ignjatovic for expressing U1 snRNP proteins and laying the foundation of this project. We thank Swiss Light Source, Daresbury and European Synchrotron Radiation Facility beamline staff, particularly C. Schulze-Bries, T. Tomizaki and A. Pauluhn at the Swiss Light Source, for their support; V. Ramakrishnan, A. Newman, A. Andreeva, A. Murzin and C. Vornheim for discussions; the current members of the Nagai group for help; S. Sengupta for support; and H. Stark for making the cryo-electron microscopy structure available. This project has been funded by the Medical Research Council and the Human Frontier Science Program (HFSP). D.A.P.K. was a recipient of a HFSP long-term fellowship. A.K.W.L. was supported by the National Science and Engineering Research Council of Canada, the ORS Fund, the Cambridge Commonwealth Trust and Sidney Sussex College Junior Research Fellowship.

**Author Information** Coordinates of the protein  $\alpha$ -carbon and the RNA phosphorus atoms have been submitted, together with structure factors, to the Protein Data Bank under accession number 3CW1. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to K.N. (kn@mrc-lmb.cam.ac.uk).

## METHODS

**Protein and RNA constructs.** Formation of ordered crystals often requires protein constructs from which flexible regions have been removed. The C-terminal sequence of U1-70K (beyond residue 216) is highly divergent and contains a long stretch of Arg-Asp/Glu and Ser-Arg repeats in vertebrates. It undergoes reversible phosphorylation<sup>46</sup> and in different states serves as a source of binding sites for numerous constitutive and alternative splicing factors<sup>13,47</sup>. Hence, the C-terminal region is likely to be unstructured and may be deleted without affecting the structural integrity of U1 snRNP. We therefore used a U1-70K construct containing the universally conserved region (residues 2–215).

The C-terminal regions of U1-C and Sm-B/Sm-B' contain multiple copies of a Pro-Pro-Pro-Gly-Met-Arg-Pro-Pro motif or its variant. It has been shown that a splicing factor, formin binding protein (WBP-4), directly binds to the C-terminal regions of U1-C and Sm-B/Sm-B' containing the proline-rich sequence<sup>44</sup>. This repetitive sequence is accessible to WBP-4 and is unlikely to be an integral part of U1 snRNP. Hence, we used U1-C (1–77) and Sm-B (1–174) lacking part of the proline-rich region for crystallization.

A U1-A-depleted U1 snRNP binds the 5' splice site as efficiently as the intact particle<sup>28</sup>. Furthermore, U1 snRNP is active in a splicing assay when nucleotides 50–91 of U1 snRNA, containing the U1-A binding site, have been deleted<sup>29</sup>. We therefore replaced the apical part of SL2 with a kissing-loop sequence<sup>30</sup> to promote crystal contacts (Supplementary Fig. 1). The functional core of U1 snRNP containing Sm-B (1–174), U1-C (1–77) and U1-70K (2–215) and the full-length constructs of Sm-D1, Sm-D2, Sm-D3, Sm-E, Sm-F and Sm-G was reconstituted with this RNA (Supplementary Fig. 1) and crystallized.

**Structure determination.** Data were integrated using MOSFLM<sup>48</sup> and scaled with SCALA<sup>49</sup> from the CCP4 suite<sup>50</sup> (Supplementary Table 1). An anomalous Patterson map of a tantalum bromide derivative was readily interpretable and heavy-atom parameters of four major and four minor sites were refined using SHARP<sup>51</sup>. A Fourier map with the tantalum bromide<sup>52</sup> phases to 7 Å resolution showed clear density for A-form RNA after solvent flipping in SOLOMON<sup>53</sup> with a 60% solvent content. The Ta<sub>6</sub>Br<sub>12</sub>, Se and Zn positions were refined together against all their respective data sets in SHARP<sup>51</sup>. The resulting map was modified in SOLOMON, as above, and simultaneously phase extended from 7.0 to 6.5 Å over 11 cycles. These phases had an overall mean figure of merit of 0.63.

**Initial model building.** Model building was carried out using O<sup>54</sup> and COOT<sup>55</sup>. The RNA helices were readily built into the U1 snRNP map. The Sm fold region of the U4 core domain structure independently solved (A.K.W.L., J.L. and K.N., unpublished results) can be fitted into the U1 map by superposing the methionine sulphur positions of the U4 core domain structure onto the Se anomalous peaks of the U1 snRNP map with a root-mean-square deviation of 1.8–2.0 Å for the four complexes, showing good agreement between these Sm ring structures. The lowest-energy NMR structure of U1-C<sup>21</sup> and a homology-modelled RBD of U1-70K<sup>56</sup> were used as initial models. The main crystal contacts include (1) base-pairing between the 5' ends of U1 snRNAs from a pair of U1 snRNPs; (2) a kissing-loop interaction between a pair of U1 snRNPs; and (3) an interaction between helix B of U1-C and SL3 of a neighbouring U1 snRNP.

**Multi-domain and multi-crystal averaging.** The four complexes in the asymmetric unit showed small but significant differences in the orientation of RNA

helices emerging from the four-way junction. The particle was divided into four domains: (1) RNA residues 1–16 and 48–134, U1-C residues 4–31, U1-70K residues 9–23, proteins Sm-D3, Sm-B, Sm-D1, Sm-D2, Sm-F, Sm-E and Sm-G; (2) SL1, U1-70K RBD and residues 63–89; (3) U1-C residues 32–61; and (4) SL4. The coordinates and matrices for each domain were used to generate masks, at a radius of 8 Å from the atoms.

An EMTS-derivative of a U1 snRNP crystal containing a Q39C mutant of U1-C diffracted to 5.5 Å resolution but was not fully isomorphous with the crystal, containing the SeMet-labelled Sm proteins, that was used for the original experimental map (Supplementary Table 1). The masks, the core SeMet crystal data, the original experimental map phases and the Q39C data were used for multi-domain, multi-crystal averaging in DMMULTI<sup>51</sup> over 30 cycles with a solvent content of 60%. The Q39C phases were refined to 5.5 Å in solvent-flattening, histogram mapping and NCS/crystal averaging modes.

The final map with an overall figure of merit of 0.62 was of excellent quality and readily interpretable. RNA helices exhibit phosphate 'bumps' in some places and some aromatic amino-acid side chains on  $\alpha$ -helices can also be identified as 'bumps' (shown using asterisks in Figs 2b and 3a). The map allowed the building of the N and C termini of Sm proteins that were not ordered in the Sm-D1–Sm-D2 and Sm-D3–Sm-B heterodimer structures<sup>22</sup>. These N- and C-terminal regions were built with the aid of heavy-atom markers as shown (Fig. 2b, c). All figures were produced using PYMOL<sup>57</sup>. The features of the U1 snRNP structure discussed in the text are common to all four complexes in the asymmetric unit. All maps shown are experimental electron density after multi-domain, multi-crystal averaging and are contoured at 1 $\sigma$ .

46. Tazi, J. *et al.* Thiophosphorylation of U1-70K protein inhibits pre-mRNA splicing. *Nature* **363**, 283–286 (1993).
47. Smith, C. W. & Valcárcel, J. Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem. Sci.* **25**, 381–388 (2000).
48. Leslie, A. G. W. The integration of macromolecular diffraction data. *Acta Crystallogr. D* **62**, 48–57 (2006).
49. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D* **62**, 72–82 (2006).
50. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
51. de la Fortelle, E. & Bricogne, G. Maximum likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–493 (1997).
52. Knäblein, J. *et al.* Ta<sub>6</sub>Br<sub>12</sub><sup>2+</sup>, a tool for phase determination of large biological assemblies by X-ray crystallography. *J. Mol. Biol.* **270**, 1–7 (1997).
53. Abrahams, J. P. & Leslie, A. G. Methods used in the structure determination of bovine mitochondrial F<sub>1</sub> ATPase. *Acta Crystallogr. D* **52**, 30–42 (1996).
54. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
55. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
56. Nagai, K., Oubridge, C., Jessen, T.-H., Li, J. & Evans, P. R. Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature* **348**, 515–520 (1990).
57. DeLano, W. L. *Pymol Molecular Viewer* (<http://www.pymol.org>) (2002).



# Accretion disk winds as the jet suppression mechanism in the microquasar GRS 1915+105

Joseph Neilsen<sup>1</sup> & Julia C. Lee<sup>1</sup>

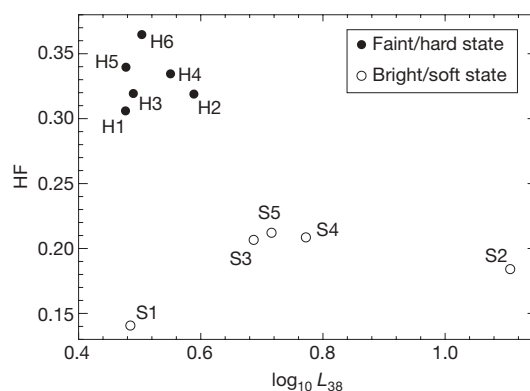
Stellar-mass black holes with relativistic jets, also known as microquasars, mimic the behaviour of quasars and active galactic nuclei<sup>1</sup>. Because timescales around stellar-mass black holes are orders of magnitude smaller than those around more distant supermassive black holes, microquasars are ideal nearby ‘laboratories’ for studying the evolution of accretion disks and jet formation in black-hole systems<sup>2</sup>. Whereas studies of black holes have revealed a complex array of accretion activity, the mechanisms that trigger and suppress jet formation remain a mystery. Here we report the presence of a broad emission line in the faint, hard states and narrow absorption lines in the bright, soft states of the microquasar GRS 1915+105. (‘Hard’ and ‘soft’ denote the character of the emitted X-rays.) Because the hard states exhibit prominent radio jets<sup>3</sup>, we argue that the broad emission line arises when the jet illuminates the inner accretion disk. The jet is weak or absent during the soft states<sup>4</sup>, and we show that the absorption lines originate when the powerful radiation field around the black hole drives a hot wind off the accretion disk<sup>5–7</sup>. Our analysis shows that this wind carries enough mass away from the disk to halt the flow of matter into the radio jet.

GRS 1915+105 is a  $14M_{\odot}$  black hole accreting matter from a  $0.8M_{\odot}$  K3 IV star in a wide 33.5-day orbit<sup>8</sup> (here  $M_{\odot}$  is the solar mass). As the first known source of superluminal jets in the Galaxy<sup>9</sup>, with a light curve exhibiting at least 14 distinct classes of high-amplitude variability due to rapid disk-jet interactions<sup>3,10–15</sup>, this microquasar provides a fascinating example of the coupling between jets and accretion disks around black holes. To study this relationship, we have analysed archival HETGS<sup>16</sup> (High Energy Transmission Grating Spectrometer) observations of GRS 1915+105 from the Chandra X-ray Observatory. Between 24 April 2000 and 14 August 2007, the HETGS observed this microquasar 11 times with high spectral resolution, probing 5 of the 14 variability classes of this enigmatic X-ray source. The data include six observations of the faint, hard, jet-producing state<sup>17</sup> (observations H1–H6) and five observations of various bright, softer states (observations S1–S5).

As one of the brightest X-ray sources in the sky, GRS 1915+105 requires Chandra observations in a special high-time-resolution mode to mitigate photon pile-up (that is, two or more photons striking a single pixel during one CCD frame time). Unfortunately, this mode is not at present well calibrated, so it is currently impossible to fit a ‘physical’ model (for example, a disk blackbody) to the broadband continuum. However, because we are mainly interested in spectral lines, we circumvent the calibration problems by fitting the X-ray continuum with a smooth tenth-order polynomial, revealing spectral features with widths less than  $0.8\text{ Å}$ . The fraction of the  $0.7\text{--}4.1\text{ Å}$  ( $3\text{--}18\text{ keV}$ ) flux emitted below  $1.4\text{ Å}$  (above  $8.6\text{ keV}$ ), which we call the hard flux fraction HF, clearly delineates the hard and soft states (Fig. 1) and the observed spectral features (Fig. 2 and Supplementary Information). We will argue that these spectral differences are strong X-ray indicators of the disk-jet coupling in GRS 1915+105.

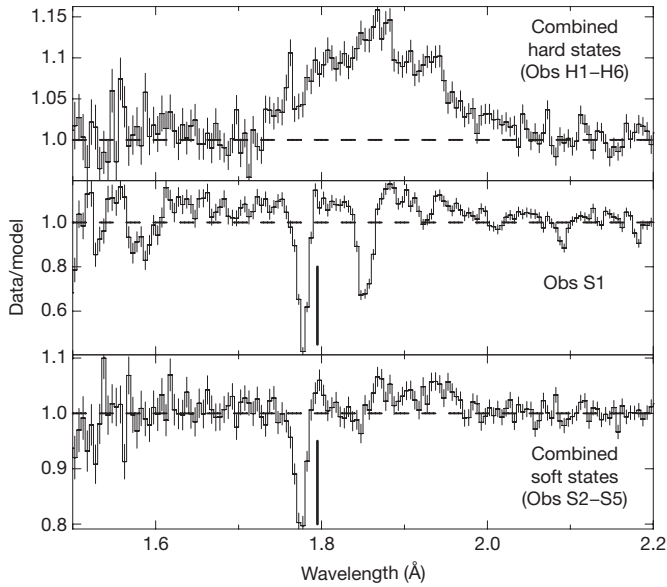
The states with high HF are dominated by a broad emission line near  $1.86\text{ Å}$  ( $6.7\text{ keV}$ ), which we tentatively identify as Fe xxv (Table 1). In contrast, strong narrow absorption lines near  $1.77\text{ Å}$  ( $7.0\text{ keV}$ ), consistent with Fe xxvi absorption blueshifted by  $\sim 1,000\text{ km s}^{-1}$ , are seen in softer states. In some cases, a weaker Fe xxv absorption line is seen at the same velocity. The absence of charge states other than hydrogen-like and helium-like iron in observations S2–S5 indicates a highly ionized absorber with ionization parameter  $\xi = L_X/nr^2$  of the order of  $10^4$  (here  $L_X$  is the X-ray luminosity,  $n$  is the gas density,  $r$  is the distance from the source of ionizing radiation, and  $\xi = 10^4$  corresponds to  $T \approx 10^6\text{ K}$ )<sup>18</sup>. These results are consistent with previous X-ray spectral studies of GRS 1915+105<sup>19–21</sup>.

It has been postulated that the accretion disk is truncated at some distance from the black hole during the low hard state<sup>3,22</sup>, when the spectrum may be dominated by hard X-rays from the corona or jet<sup>4,23</sup>. Coupled with infrared studies implicating jet activity in the production of emission lines in GRS 1915+105<sup>24</sup>, we conclude that the broad Fe xxv emission line is produced when the inner edge of the disk is illuminated by these hard X-rays. Our interpretation is substantiated by the fact that the equivalent width of the Fe xxv emission line increases with both  $L_X$  and the radio flux at  $15\text{ GHz}$  ( $S_{15\text{ GHz}}$ ) as measured by the Ryle telescope<sup>15</sup> (Fig. 3). Under the assumption that the line broadening ( $\geq 12,500\text{ km s}^{-1}$ ) is due to orbital motion in a



**Figure 1 | The X-ray luminosity and hard flux fraction for the 11 archival HETGS observations of GRS 1915+105.**  $L_{38}$  is the X-ray luminosity in units of  $10^{38}\text{ erg s}^{-1}$ , measured with RXTE (Rossi X-ray Timing Explorer) from  $0.7\text{--}4.1\text{ Å}$  ( $3\text{--}18\text{ keV}$ ), assuming a distance of  $12.5\text{ kpc}$  (ref. 9) and neutral hydrogen absorption ( $N_H = 5 \times 10^{22}\text{ cm}^{-2}$ )<sup>19</sup> along the line of sight. The hard flux fraction (HF), used as a proxy for the strength of the Comptonized emission from the corona or jet, is defined as the ratio of the unabsorbed continuum flux from  $0.7\text{--}1.4\text{ Å}$  ( $8.6\text{--}18\text{ keV}$ ) to  $0.7\text{--}4.1\text{ Å}$  ( $3\text{--}18\text{ keV}$ ). The 11 observations are classified as hard or soft based on previous X-ray studies<sup>10</sup>; as expected, the hard states have a higher HF.

<sup>1</sup>Harvard University Department of Astronomy, 60 Garden Street, MS-10, Cambridge, Massachusetts 02138, USA.



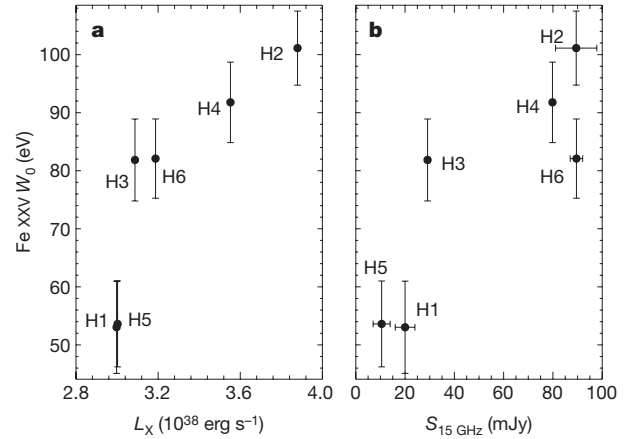
**Figure 2 | The data/model ratio for the continuum fits to the HETGS observations of GRS 1915+105.** We plot the combined hard states (observations (Obs) H1–H6) and combined soft states (observations S2–S5) for clarity; observation S1 is shown alone to highlight other strong lines. Plots of the individual spectra can be found in Supplementary Information. The broad Fe xxv emission line distinguishes periods of significant hard X-ray illumination from softer states, which are dominated by strong Fe xxvi absorption lines. We use a simple Gaussian to measure the properties of the broad Fe xxv line (with the assumption that it is not a blend of emission from several Fe charge states) and find it has a line width  $\geq 12,500 \text{ km s}^{-1}$ , much larger than the orbital velocity of either the companion star or the black hole<sup>8</sup>. We note that it represents a lower limit, because our continuum-fitting procedure would mask the red wing of any relativistically broadened emission line; the implied inner disk radius of  $255R_S$  is thus an upper limit. While the emission line width probes the size of the disk, the P Cygni profile of the absorption lines constrains the geometry of the accretion disk wind. The vertical line indicates the position of the weak P Cygni emission component. Because this emission component is weak, the wind must be confined to the equatorial plane of the disk. As the binary is viewed at an inclination  $i = 70^\circ$  (ref. 9), we can suppose that  $i < 20^\circ$  above the midplane of the disk is a reasonable estimate (implying  $f < 5\%$ ). Spectral analysis was performed with the ISIS<sup>28</sup> spectral fitting package. The errors shown are 68% confidence limits on the data/model ratio.

Keplerian disk, the line is emitted at  $r \leq 1.1 \times 10^9 \text{ cm}$  ( $\leq 255R_S$ , where  $R_S$  is the Schwarzschild radius of the black hole). This is a reasonable upper limit for the inner edge of the truncated disk.

**Table 1 | Spectral properties of GRS 1915+105**

Obs no.	Obs ID	X-ray state <sup>10</sup>	$L_{38}$	HF	$S_{15 \text{ GHz}}$ (mJy)	Line ID	$\lambda_0$ (Å)	$\lambda_{\text{obs}}$ (Å)	$\Delta v$ (km s <sup>-1</sup> )	$W_0$ (eV)	$\sigma$ (km s <sup>-1</sup> )
S1	7485	$\phi$	3.1	0.14	...	Fe xxvi	1.7807	$1.7774^{+0.0004}_{-0.0005}$	$-560^{+70}_{-80}$	$-29.9^{+1.3}_{-1.6}$	$650^{+150}_{-160}$
S2	6581	$\gamma$	12.8	0.18	$5 \pm 3$	Fe xxvi	1.7807	$1.775 \pm 0.001$	$-1,000^{+240}_{-220}$	$-21.9^{+2.2}_{-2.7}$	$1,160^{+280}_{-250}$
S3	1945	$\rho$	4.9	0.21	$3 \pm 2$	Fe xxvi	1.7807	$1.772 \pm 0.002$	$-1,420^{+320}_{-310}$	$-7.2 \pm 1.7$	$<1,120$
S4	6580	$\beta$	5.9	0.21	20–60	Fe xxvi	1.7807	$1.774 \pm 0.002$	$-1,100^{+360}_{-300}$	$-19.3^{+3.2}_{-3.5}$	$980^{+450}_{-420}$
S5	6579	$\beta$	5.2	0.21	20–60	Fe xxvi	1.7807	$1.775^{+0.002}_{-0.003}$	$-910^{+390}_{-430}$	$-13.3^{+3.0}_{-2.9}$	$<1,080$
H1	660	$\chi$	3.0	0.31	$20 \pm 4^{19}$	Fe xxvi	1.7807	$1.775^{+0.004}_{-0.003}$	$-910^{+630}_{-570}$	$-4.8^{+1.6}_{-2.5}$	$<1,300$
H1	660	$\chi$	3.0	0.31	$20 \pm 4^{19}$	Fe xxv	1.868	$1.89 \pm 0.02$	$2,830^{+3,830}_{-3,840}$	$53.0 \pm 7.9$	$17,820^{+2,000}_{-2,950}$
H2	4587	$\chi$	3.9	0.32	$89 \pm 8$	Fe xxv	1.868	$1.865 \pm 0.006$	$460 \pm 980$	$101.1 \pm 6.4$	$11,120^{+840}_{-750}$
H3	1944	$\chi$	3.1	0.32	$29 \pm 1$	Fe xxv	1.868	$1.867 \pm 0.009$	$-260^{+1,460}_{-1,480}$	$81.9 \pm 7.1$	$12,210^{+1,290}_{-1,100}$
H4	4589	$\chi$	3.6	0.33	$80 \pm 1$	Fe xxv	1.868	$1.861 \pm 0.007$	$-1,230^{+1,180}_{-1,150}$	$91.8 \pm 6.9$	$11,130^{+1,110}_{-960}$
H5	1946	$\chi$	3.0	0.34	$10 \pm 3$	Fe xxv	1.868	$1.86^{+0.01}_{-0.02}$	$-2,000^{+2,300}_{-2,340}$	$53.6 \pm 7.4$	$11,950^{+1,930}_{-1,580}$
H6	4588	$\chi$	3.2	0.36	$90 \pm 3$	Fe xxv	1.868	$1.867 \pm 0.008$	$-130^{+1,230}_{-1,240}$	$82.1 \pm 6.8$	$10,590^{+1,080}_{-950}$

The 11 HETGS observations are listed in order of increasing hard flux fraction with their Chandra observation ID numbers and relevant spectral properties. ‘S’ indicates a soft state and ‘H’ indicates a hard state. The Greek letters identify the observations with one of the 14 variability classes of this microquasar<sup>10</sup>.  $L_{38}$  is the 0.7–4.1 Å (3–18 keV) luminosity, measured with RXTE, in units of  $10^{38} \text{ erg s}^{-1}$ . HF is the hard flux fraction, defined as the ratio of the unabsorbed 0.7–1.4 Å (8.6–18 keV) to 0.7–4.1 Å (3–18 keV) continuum flux.  $S_{15 \text{ GHz}}$  is the radio flux at 15 GHz, measured by the Ryle telescope<sup>15</sup>. The line ID is our identification of the strongest line; we detect Fe xxvi in absorption and Fe xxv in emission.  $\lambda_0$  is the ion’s laboratory wavelength and  $\lambda_{\text{obs}}$  is the observed wavelength.  $\Delta v$  is the corresponding Doppler velocity.  $W_0$  is the line equivalent width, and  $\sigma$  is the line width. All errors correspond to 90% confidence limits. Obs H1 exhibits both a weak broad emission line and a weak narrow absorption line because of its intermediate hard flux fraction.

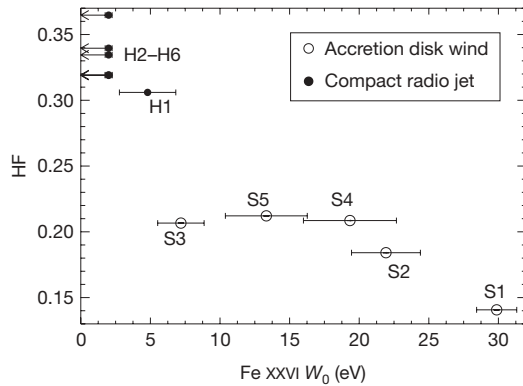


**Figure 3 | The equivalent width of the broad Fe xxv emission line in the hard states of GRS 1915+105 as a function of X-ray luminosity and radio flux.** **a**, X-ray luminosity,  $L_X$ ; **b**, radio flux at 15 GHz,  $S_{15 \text{ GHz}}$ . Because the emission line equivalent width,  $W_0$ , is correlated with both  $L_X$  and  $S_{15 \text{ GHz}}$ , it is more likely that the accretion disk is illuminated by the X-ray emitting base of the radio jet than by a hot, tenuous corona. Because HF does not obviously scale with  $L_X$ , more detailed studies are required to determine the complex relationship between the X-ray illumination that causes the broad lines and the hard X-rays that ionize the wind.  $W_0$  was measured on the basis of simple Gaussian fits; the actual values will be larger if continuum uncertainties have masked the broad red wing of the line. Errors shown for  $W_0$  correspond to 90% confidence limits on the emission line flux; errors shown for  $S_{15 \text{ GHz}}$  (ref. 15) are also 90% confidence limits.

In comparison, the inner edge of the accretion disk may lie much closer to the black hole (as little as  $r = 3R_S$  or less) during bright soft states<sup>25</sup>. In this context, we suggest that the absorption lines seen in GRS 1915+105 originate in an accretion disk wind. Several lines of evidence support this interpretation: the absorption lines (1) only appear during softer states, when the disk may be prominent and hard X-ray illumination is relatively weak (Fig. 4), (2) are all narrow and blueshifted (see Table 1), implying material flowing into our line of sight, and are accompanied by (3) a weak emission line at a slightly longer wavelength (Fig. 2). Together, (2) and (3) constitute a P Cygni profile, a classic wind signature. As a K-type companion cannot drive a strong wind, the wind must originate in the accretion disk. Furthermore, the wind speed corresponds to the escape velocity from the black hole at a distance of  $r = 2.5 \times 10^{11} \text{ cm}$  ( $=53,460R_S$ ), which is well inside the accretion disk of GRS 1915+105.

Given the high luminosity of this black-hole binary and the strong variation of the equivalent width of the Fe xxvi absorption line with  $L_X$





**Figure 4 | The hard flux fraction and the equivalent width of the Fe xxvi absorption line seen in GRS 1915+105.** The hard flux fraction (HF) is defined as the ratio of the unabsorbed 0.7–1.4 Å (8.6–18 keV) to 0.7–4.1 Å (3–18 keV) continuum fluxes, measured with RXTE. This figure shows HF decreasing with  $W_0$ , implying that the jet weakens as the wind strength increases (and vice versa). It appears that by carrying a significant amount of matter away from the accretion disk, strong winds can suppress jet production. This figure illustrates the nature of the complex competition between the wind and the jet, because the trend can also be understood in terms of the hard X-ray illumination of the wind. In the hardest states, the corona/jet may completely photoionize the wind, so that the gas is transparent, and therefore the absorption lines are weak or absent and the mass loss rate in the wind decreases significantly. This effect can also explain (for higher hard flux fraction) the disappearance of the lower-ionization absorption lines present in observation S1. Because of the complicated coupling between the disk and jet, we do not rule out broadened emission lines in softer states, or narrow absorption lines in harder states. In fact, the presence of both a broad emission and narrow absorption lines in observation H1 provides further evidence of active competition between the wind and the jet. These results imply that HF is a viable indicator of accretion physics around black holes. The errors shown for the equivalent width correspond to 90% confidence limits on the absorbed line flux.

(Table 1), it is likely that radiation pressure plays a role in driving this highly ionized wind<sup>7</sup>. However, radiation pressure alone (mainly imparted by ultraviolet emission lines) is inefficient at transferring momentum to a wind at  $\xi > 10^3$  (ref. 26). But at the high luminosity of GRS 1915+105, X-ray heating and thermal pressure can provide the extra boost to drive a hot, fast wind off the accretion disk for  $r > 0.01 r_c$ , where the critical radius  $r_c$  is given by  $r_c = (9.8 \times 10^9) \times (M_{\text{BH}}/M_\odot) / T_{\text{C8}} \text{ cm}$ . Here  $M_{\text{BH}}$  is the black-hole mass, and  $T_{\text{C8}}$  is the electron temperature in units of  $10^8 \text{ K}$  (refs 5, 6). As our analysis indicates that the wind temperature  $T_{\text{wind}} \approx 10^6 \text{ K}$ , the wind could originate at any  $r > 1.4 \times 10^{11} \text{ cm}$  ( $\approx 29,340 R_g$ ). Although the relevant electron temperature could be much higher<sup>5</sup>, this launching radius is consistent with our earlier estimate from the blueshift of the wind, so that thermal driving assisted by radiation pressure successfully explains the origin of this wind.

To estimate the mass loss rate from the wind, we fit the spectra with a photoionized absorption model that calculates the ionization balance for a shell of gas surrounding a central X-ray source<sup>18</sup>. This model is characterized mainly by the absorbing gas column density and ionization parameter  $\xi \approx 10^4$ ; for a fixed luminosity and wind speed, a higher ionization parameter implies a smaller mass loss rate<sup>19</sup>. Assuming an accretion efficiency  $\eta$ , accretion rate  $\dot{M}_{\text{acc}}$ , and wind covering factor  $f$ , our model implies a wind mass loss rate of  $\dot{M}_w = 188 \dot{M}_{\text{acc}} \eta f$ . With  $f < 5\%$  (Fig. 2 legend) and  $\eta = 6\%$  (ref. 27), we calculate  $\dot{M}_w < 0.59 \dot{M}_{\text{acc}}$  ( $\sim 10^{-8} M_\odot \text{ yr}^{-1}$ ).

Interestingly, this wind drives approximately the same mass loss as the radio jet<sup>17</sup>, suggesting that GRS 1915+105 is able to maintain a rough equilibrium between mass accretion and outflow, independent of its spectral state and the outflow mechanism, over the span of our observations. Furthermore, the noticeable decrease of HF with the equivalent width of the absorption lines (Fig. 4) indicates a complex

competition between the accretion disk wind and the radio jet. When HF decreases, there are fewer hard X-rays available to over-ionize the wind, allowing it to carry away more of the matter that sustains the jet. Thus it appears that Comptonization and photoionization mediate the coupling between the jet and the disk in GRS 1915+105.

This is a strong indication that like their supermassive counterparts, stellar-mass black holes can regulate their accretion rate by feedback into their environments. More importantly, these observations clearly demonstrate that at sufficiently high luminosities in GRS 1915+105, the intense radiation field of the disk redirects the accretion flow, away from the radio jet and into a wind. By revealing a surprisingly simple jet-quenching mechanism in GRS 1915+105, our results point to fundamental new insights into the long-term disk–jet coupling around accreting black holes and hint at tantalizing evidence of the mechanism by which stellar-mass black holes regulate their own growth.

Received 5 September 2008; accepted 5 February 2009.

1. Mirabel, I. F. & Rodríguez, L. F. Sources of relativistic jets in the Galaxy. *Annu. Rev. Astron. Astrophys.* **37**, 409–443 (1999).
2. Fender, R. P. & Belloni, T. GRS 1915+105 and the disc-jet coupling in accreting black hole systems. *Annu. Rev. Astron. Astrophys.* **42**, 317–364 (2004).
3. Klein-Wolt, M. et al. Hard X-ray states and radio emission in GRS 1915+105. *Mon. Not. R. Astron. Soc.* **331**, 745–764 (2002).
4. Fender, R. et al. Quenching of the radio jet during the X-ray high state of GX 339–4. *Astrophys. J.* **519**, L165–L168 (1999).
5. Begelman, M. C., McKee, C. F. & Shields, G. A. Compton heated winds and coronae above accretion disks. I Dynamics. *Astrophys. J.* **271**, 70–88 (1983).
6. Proga, D. & Kallman, T. R. On the role of the ultraviolet and X-ray radiation in driving a disk wind in X-ray binaries. *Astrophys. J.* **565**, 455–470 (2002).
7. Proga, D. Winds from accretion disks driven by radiation and magnetocentrifugal force. *Astrophys. J.* **538**, 684–690 (2000).
8. Greiner, J., Cuby, J. G. & McCaughrean, M. J. An unusually massive stellar black hole in the Galaxy. *Nature* **414**, 522–524 (2001).
9. Mirabel, I. F. & Rodríguez, L. F. A superluminal source in the Galaxy. *Nature* **371**, 46–48 (1994).
10. Belloni, T., Klein-Wolt, M., Mendez, M., van der Klis, M. & van Paradijs, J. A model-independent analysis of the variability of GRS 1915+105. *Astron. Astrophys.* **355**, 271–290 (2000).
11. Hannikainen, D. et al. Characterizing a new class of variability in GRS 1915+105 with simultaneous INTEGRAL/RXTE observations. *Astron. Astrophys.* **435**, 995–1004 (2005).
12. Mirabel, I. F. et al. Accretion instabilities and jet formation in GRS 1915+105. *Astron. Astrophys.* **330**, L9–L12 (1998).
13. Eikenberry, S. S., Matthews, K., Morgan, E. H., Remillard, R. & Nelson, R. W. Evidence for a disk-jet interaction in the microquasar GRS 1915+105. *Astrophys. J.* **494**, L61–L64 (1998).
14. Fender, R. P., Pooley, G. G., Brocksopp, C. & Newell, S. J. Rapid infrared flares in GRS 1915+105: Evidence for infrared synchrotron emission. *Mon. Not. R. Astron. Soc.* **290**, L65–L69 (1997).
15. Pooley, G. G. & Fender, R. P. The variable radio emission from GRS 1915+105. *Mon. Not. R. Astron. Soc.* **292**, 925–933 (1997).
16. Canizares, C. et al. The Chandra high-energy transmission grating: Design, fabrication, ground calibration, and 5 years in flight. *Publ. Astron. Soc. Pacif.* **117**, 1144–1171 (2005).
17. Dhawan, V., Mirabel, I. F. & Rodríguez, L. F. AU-scale synchrotron jets and superluminal ejecta in GRS 1915+105. *Astrophys. J.* **543**, 373–385 (2000).
18. Kallman, T. R. & Bautista, M. Photoionization and high-density gas. *Astrophys. J.* **133** (Suppl.), 221–253 (2001).
19. Lee, J. C. et al. High-resolution Chandra HETGS and Rossi X-Ray Timing Explorer observations of GRS 1915+105: A hot disk atmosphere and cold gas enriched in iron and silicon. *Astrophys. J.* **567**, 1102–1111 (2002).
20. Munro, M., Morgan, E. H. & Remillard, R. Quasi-periodic oscillations and spectral states in GRS 1915+105. *Astrophys. J.* **527**, 321–340 (1999).
21. Kotani, T. et al. ASCA observations of the absorption line features from the superluminal jet source GRS 1915+105. *Astrophys. J.* **539**, 413–423 (2000).
22. Esin, A. A., McClintock, J. E. & Narayan, R. Advection-dominated accretion and the spectral states of black hole X-ray binaries: Application to Nova Muscae 1991. *Astrophys. J.* **489**, 865–889 (1997).
23. Markoff, S., Nowak, M. A. & Wilms, J. Going with the flow: Can the base of jets subsume the role of compact accretion disk coronae? *Astrophys. J.* **635**, 1203–1216 (2005).
24. Eikenberry, S. et al. Spectroscopy of infrared flares from the microquasar GRS 1915+105. *Astrophys. J.* **506**, L31–L34 (1998).
25. McClintock, J. et al. The spin of the near-extreme Kerr black hole GRS 1915+105. *Astrophys. J.* **652**, 518–539 (2006).
26. Proga, D., Stone, J. M. & Kallman, T. R. Dynamics of line-driven winds in active galactic nuclei. *Astrophys. J.* **543**, 686–696 (2000).

27. Janiuk, A., Czerny, B. & Siemiginowska, A. Radiation pressure instability as a variability mechanism in the microquasar GRS 1915+105. *Astrophys. J.* **542**, L33–L36 (2000).
28. Houck, J. C. & Denicola, L. A. in *Astronomical Data Analysis Software and Systems IX* (eds Manset, N., Veillet, C. & Crabtree, D.) 591–594 (ASP Conference Series Vol. 216, 2000).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We acknowledge support from the Harvard University Graduate School of Arts and Sciences (J.N.) and the Faculty of Arts and Sciences

(J.C.L.). We thank G. Pooley for providing the radio data used in this paper and we acknowledge conversations with R. Remillard, who provided the Rossi X-ray Timing Explorer spectra, and M. Begelman.

**Author Contributions** J.N. processed the data, performed spectral analysis, and wrote the paper. J.C.L. commented extensively on the manuscript. Both authors discussed the results at length.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to J.N. ([jneilsen@cfa.harvard.edu](mailto:jneilsen@cfa.harvard.edu)).

# The impact and recovery of asteroid 2008 TC<sub>3</sub>

P. Jenniskens<sup>1</sup>, M. H. Shaddad<sup>2</sup>, D. Numan<sup>2</sup>, S. Elsir<sup>3</sup>, A. M. Kudoda<sup>2</sup>, M. E. Zolensky<sup>4</sup>, L. Le<sup>4,5</sup>, G. A. Robinson<sup>4,5</sup>, J. M. Friedrich<sup>6,7</sup>, D. Rumble<sup>8</sup>, A. Steele<sup>8</sup>, S. R. Chesley<sup>9</sup>, A. Fitzsimmons<sup>10</sup>, S. Duddy<sup>10</sup>, H. H. Hsieh<sup>10</sup>, G. Ramsay<sup>11</sup>, P. G. Brown<sup>12</sup>, W. N. Edwards<sup>12</sup>, E. Tagliaferri<sup>13</sup>, M. B. Boslough<sup>14</sup>, R. E. Spalding<sup>14</sup>, R. Dantowitz<sup>15</sup>, M. Kozubal<sup>15</sup>, P. Pravec<sup>16</sup>, J. Borovicka<sup>16</sup>, Z. Charvat<sup>17</sup>, J. Vaubillon<sup>18</sup>, J. Kuiper<sup>19</sup>, J. Albers<sup>1</sup>, J. L. Bishop<sup>1</sup>, R. L. Mancinelli<sup>1</sup>, S. A. Sandford<sup>20</sup>, S. N. Milam<sup>20</sup>, M. Nuevo<sup>20</sup> & S. P. Worden<sup>20</sup>

In the absence of a firm link between individual meteorites and their asteroidal parent bodies, asteroids are typically characterized only by their light reflection properties, and grouped accordingly into classes<sup>1–3</sup>. On 6 October 2008, a small asteroid was discovered with a flat reflectance spectrum in the 554–995 nm wavelength range, and designated 2008 TC<sub>3</sub> (refs 4–6). It subsequently hit the Earth. Because it exploded at 37 km altitude, no macroscopic fragments were expected to survive. Here we report that a dedicated search along the approach trajectory recovered 47 meteorites, fragments of a single body named Almahata Sitta, with a total mass of 3.95 kg. Analysis of one of these meteorites shows it to be an achondrite, a polymict ureilite, anomalous in its class: ultra-fine-grained and porous, with large carbonaceous grains. The combined asteroid and meteorite reflectance spectra identify the asteroid as F class<sup>3</sup>, now firmly linked to dark carbon-rich anomalous ureilites, a material so fragile it was not previously represented in meteorite collections.

The asteroid was discovered by the automated Catalina Sky Survey telescope at Mount Lemmon, Arizona on October 6 06:39 UTC (ref. 4). Early orbital solutions showed an impact 19 h after discovery with a predicted impact location in the Nubian Desert of northern Sudan<sup>5,6</sup> (Table 1). Numerous astronomical observatories imaged the object until it entered the Earth's umbra on October 7 01:49 UTC. In the previous two hours, its brightness oscillated with an amplitude of 1.02 mag at main periods of  $49.0338 \pm 0.0007$  s and  $96.987 \pm 0.003$  s, and their harmonics, revealing that the asteroid was in a non-principal-axis rotation state<sup>7</sup>. The oscillation was centred on absolute visible

magnitude  $H = 30.9 \pm 0.1$  (using a phase angle slope parameter  $G = 0.15$ ). This is a measure of the asteroid's size.

Eyewitnesses in Wadi Halfa and at Station 6 (a train station between Wadi Halfa and Al Khurtum, Sudan) in the Nubian Desert described a rocket-like fireball with an abrupt ending. Sensors aboard US government satellites first detected the bolide at 65 km altitude at 02:45:40 UTC (ref. 8). The optical signal consisted of three peaks spanning 3.5 s, with most of the radiated energy in the middle 1-s pulse at an inferred altitude of about 37 km, and a final pulse 1 s later. Meteosat 8 (ref. 9) detected the brightest optical signal when the asteroid was at  $37.5 \pm 1.0$  km. Rapidly fading infrared radiation was detectable down until at least  $32.7 \pm 0.7$  km. The 10- $\mu$ m Si–O band of glowing dust was the dominant feature in a seven-channel 6–13  $\mu$ m infrared spectrum taken  $\sim 1$  s after the explosion. The height of the dust cloud was  $35.7 \pm 0.7$  km. Independently, we measured this altitude at 35–42 km, with no significant dust deposition below 33 km, based on UK Meteorological Office<sup>10</sup> wind model data and ground-based images of the lingering train<sup>11</sup> taken from Wadi Halfa at sunrise (03:22–03:27 UTC).

Unexpectedly, some meteorites survived the explosion. Fifteen fresh-looking meteorites with a total mass of 563 g were recovered by 45 students and staff of the University of Khartoum during a field campaign on 5–8 December 2008. A second search on 25–30 December (72 participants) raised the total to 47 meteorites and 3.95 kg. Masses range from 1.5 g to 283 g, spread for 29 km along the approach path in a manner expected for debris from 2008 TC<sub>3</sub> (Fig. 1).

Nearly all recovered meteorites show a broken face with no corresponding pieces nearby (Fig. 2). One intact fully crusted meteorite was perfectly oriented in flight, with only a single side exposed to the oncoming air stream and one rotational degree of freedom (Fig. 2e), suggesting that this secondary fragmentation was caused by centrifugal forces or uneven dynamic pressure from rapid tumbling.

Almahata Sitta is a fine-grained fragmental breccia. A small 1.5 g meteorite (no. 7) was broken under 35 lb peak pressure to create a fresh surface for analysis—all results reported here are from this meteorite. It had a tensile strength of only  $56 \pm 26$  MPa, cracking along a white layer, rich in pyroxene, sprinkled with darker areas rich in carbonaceous matter (Fig. 2a).

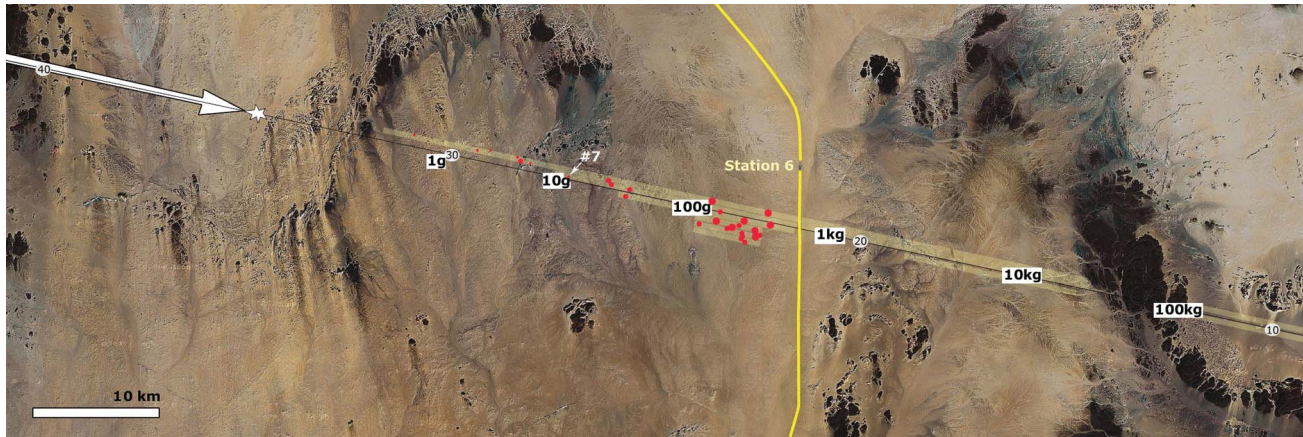
**Table 1 | Orbital parameters of 2008 TC<sub>3</sub> used to calculate the approach path**

Symbol	Parameter	Value
$a$	Semimajor axis	$1.308201 \pm 0.000009$ AU
$q$	Perihelion distance	$0.899957 \pm 0.000002$ AU
$\omega$	Argument of perihelion	$234.44897 \pm 0.00008^\circ$
$\Omega$	Longitude of ascending node	$194.101138 \pm 0.000002^\circ$
$i$	Inclination	$2.54220 \pm 0.00004^\circ$
$T_p$	Perihelion time	2008 November 20.3989 $\pm$ 0.0001 UT

These parameters are JPL solution 15; equinox J2000, 2008 October 07.0 TDB or Barycentric Dynamical Time. The astrometric position of 295 observations were used. This ephemeris, when projected to an altitude of 50 km, predicts an entry velocity of  $12.42 \text{ km s}^{-1}$  at a shallow  $20^\circ$  angle relative to the surface, with a perpendicular uncertainty in position of only  $\pm 100$  m.

<sup>1</sup>SETI Institute, Carl Sagan Center, 515 North Whisman Road, Mountain View, California 94043, USA. <sup>2</sup>Physics Department, University of Khartoum, PO Box 321, Khartoum 11115, Sudan. <sup>3</sup>Physics Department, Juba University, Juba, Sudan. <sup>4</sup>NASA Johnson Space Center, Mail Code KT, Houston, Texas 77058, USA. <sup>5</sup>Jacobs Technologies Engineering Science Contact Group (ESCG), Johnson Space Center, Houston, Texas 77058, USA. <sup>6</sup>Department of Chemistry, Fordham University, 441 East Fordham Road, Bronx, New York 10458, USA. <sup>7</sup>Department of Earth and Planetary Sciences, American Museum of Natural History, 79th Street at Central Park West, New York, New York 10024, USA. <sup>8</sup>Geophysical Laboratory, Carnegie Institution of Washington, 5251 Broad Branch Road, NW, Washington DC 20015-1305, USA. <sup>9</sup>Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA. <sup>10</sup>School of Mathematics and Physics, Queen's University Belfast, University Road, Belfast BT7 1NN, UK. <sup>11</sup>Armagh Observatory, College Hill, Armagh BT61 9DG, UK. <sup>12</sup>Department of Physics and Astronomy, University of Western Ontario, London, Ontario N6A 3K7, Canada. <sup>13</sup>ET Space Systems, 5990 Worth Way, Camarillo, California 93012, USA. <sup>14</sup>Sandia National Laboratories, PO Box 5800, Albuquerque, New Mexico 87185, USA. <sup>15</sup>Clay Center Observatory, Dexter and Southfield Schools, 20 Newton Street, Brookline, Massachusetts 02445, USA. <sup>16</sup>Astronomical Institute of the Academy of Sciences, Fričova 298, 25165 Ondřejov Observatory, Czech Republic. <sup>17</sup>Czech Hydrometeorological Institute, Na Šabatce 17, 143 06 Praha 4, Czech Republic. <sup>18</sup>Institut de Mécanique Céleste et de Calcul des Éphémérides, 77 avenue Denfert-Rochereau, 75014 Paris, France. <sup>19</sup>Dutch Meteor Society, Akker 141, 3732 XD De Bilt, The Netherlands. <sup>20</sup>NASA Ames Research Center, Mail Stop 245-6, Moffett Field, California 94035, USA.





**Figure 1** | Map of the Nubian Desert of northern Sudan with the ground-projected approach path of the asteroid and the location of the recovered meteorites. 2008 TC<sub>3</sub> moved from a geodetic longitude of 31.80381° E and latitude of +20.85787° N at 50 km altitude, to 32.58481° E, +20.70569° N at 20 km altitude above the WGS-84 ellipsoid. White arrow represents the path of the 2008 TC<sub>3</sub> fireball with the projected, non-decelerating ground path represented as a thin black line (altitude labels in km, within white ovals). The sizes of the red symbols indicate small (1–10 g), medium (10–100 g) and large (100–1,000 g) meteorites. Our dark-flight calculations show that 270-g

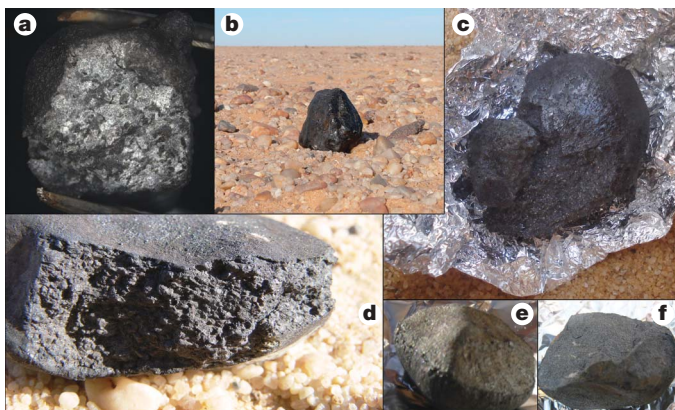
fragments would have stopped ablating at around 32 km altitude, falling vertically on the ground at 30–60 m s<sup>-1</sup>. Labels in white rectangles mark the position where meteorites of indicated masses are predicted to have fallen (calculations assume spheres released at 12.4 km s<sup>-1</sup> from detonation at 37 km altitude, white star). In light yellow is shown the area that was systematically searched. Special attention was given to possible large fragments further down track, but none were found. Such larger masses would have carried residual forward velocity. The yellow line marks the path of the local train tracks with the location of Station 6 labelled.

Classification of the meteorite was based on oxygen isotopes, bulk chemistry, and mineralogy. The oxygen isotope abundance of two fragments was measured as:  $\Delta^{17}\text{O} = -0.147\text{‰}$  and  $-0.501\text{‰}$ ,  $\delta^{17}\text{O} = 3.90\text{‰}$  and  $3.56\text{‰}$ , and  $\delta^{18}\text{O} = 7.70\text{‰}$  and  $7.72\text{‰}$  relative to Standard Mean Ocean Water (SMOW). A third sample, in contact with fusion crust, gave  $\Delta^{17}\text{O} = -0.539\text{‰}$ ,  $\delta^{17}\text{O} = 3.09\text{‰}$ , and  $\delta^{18}\text{O} = 6.89\text{‰}$  SMOW. These values scatter along the carbonaceous chondrite anhydrous mineral (CCAM) slope of  $\delta^{17}\text{O}/\delta^{18}\text{O}$ , on the upper edge of the compositional field of ureilites<sup>12</sup>—see Supplementary Information. Bulk chemistry shows that trace element abundances are achondritic (tabulated in Supplementary Information). Rare earth element (REE) abundances relative to CI chondrites steadily increase with atomic number from 0.1 to 0.6 CI, and possess a distinct negative Eu anomaly, closely resembling the bulk analyses of many ureilites, and generally interpreted as indicative

of the material being a residue of partial melting<sup>13,14</sup>. The relatively high ( $\geq 0.1$  CI) REE abundances in Almahata Sitta are consistent with it being a polymict ureilite, which as a group have higher REE concentrations than the more common monomict ureilites<sup>14</sup>. (The ‘polymict’ modifier refers to the presence of olivine and pyroxene-rich lithic clasts among ureilitic clasts.) The sample has subrounded mineral fragments and fine-grained olivine aggregates embedded in a cataclastic matrix of ureilitic material (Fig. 3A, B)<sup>15</sup>. Only one case of zoned olivine was found. Shock effects are not apparent. The examined samples have considerable porosity, ranging from 10% to 25%; the walls of pores are commonly coated by anhedral to euhedral crystals of low-calcium pyroxene (Fs<sub>2</sub>Wo<sub>3</sub>) and olivine (Fa<sub>12–14</sub>), and in some instances spherules of kamacite and botryoidal masses of Cr-bearing troilite (Fig. 3C). (Here Fs indicates ferrosilite, Wo wollastonite, Fa fayalite, and Fa<sub>12–14</sub> indicates 12–14% of this component.) These could be vapour deposits. Aggregates of carbonaceous material, up to 0.5 mm in size, are common and primarily consist of fine-grained graphite, making the rock dark. Some diamond and aliphatic carbon is also present (Fig. 3D).

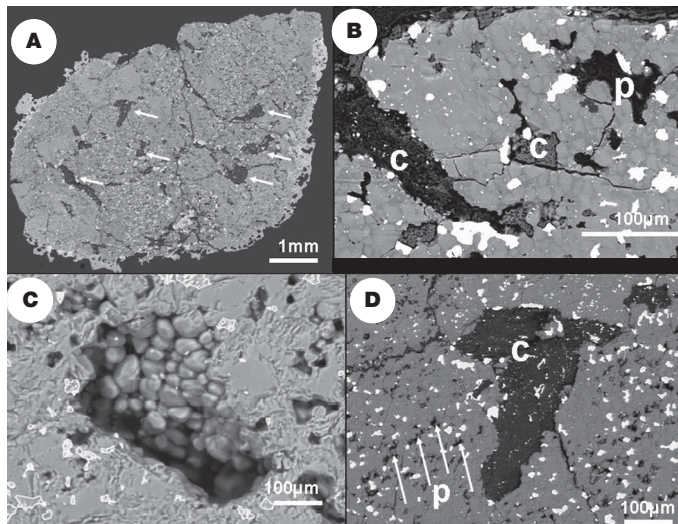
On the basis of the above information, Almahata Sitta is classified as an anomalous polymict ureilite<sup>14,16</sup>. Ureilites are coarse-grained, ultramafic rocks believed to be either magmatic cumulates or partial melt residues. Mineral compositions of Almahata Sitta are not anomalous, but the textures are, including rare zoning of olivine, larger size carbonaceous aggregates, fine-grained texture, high metal content, and high porosity with possible vapour-phase mineral growth of olivine (consistent with rapid cooling of an impact-produced melt). Other ureilites have a bulk density of  $3.05 \pm 0.22 \text{ g cm}^{-3}$  and an average micro-porosity of 9% (range 6–20%)<sup>17</sup>. The bulk density of Almahata Sitta varies from fragment to fragment. The most precisely measured values (in g cm<sup>-3</sup>) are  $2.10 \pm 0.06$  (no. 14, 152.6 g) and  $2.50 \pm 0.08$  (no. 16, 171.1 g). Assuming an average ureilite grain density<sup>17</sup> of  $3.35 \text{ g cm}^{-3}$ , this puts the porosity of Almahata Sitta in the 25–37% range, equal to the high porosities of primitive carbonaceous chondrite meteorites<sup>17</sup>.

The recovered meteorites represent only  $\sim 0.005\%$  of the initial mass, derived as follows: most are darker than the fractured surface of no. 7 (Fig. 2). Using the V-band albedo of  $0.046 \pm 0.005$ , measured for the dark phase of the meteorite, the asteroid’s absolute visual magnitude translates to an asteroid diameter of  $4.1 \pm 0.3 \text{ m}$  (ref. 18). If the density were  $2.3 \pm 0.2 \text{ g cm}^{-3}$ , then the pre-atmospheric mass was



**Figure 2** | Macroscopic features of the Almahata Sitta meteorite. **a**, Evidence of clasts in meteorite no. 7 (1 cm diameter) in a fresh fracture surface induced by pressure in the laboratory. **b**, Meteorite no. 15 (4 cm diameter), *in situ*, shows rounded shape of ablated surface. **c**, Meteorite no. 4 (14 g), placed on aluminium foil, shows the dark interior of a surface fractured upon impact. **d**, Meteorite no. 14 (2 × 7 cm), *in situ*, shows millimetre-sized grains in a weathered surface that was broken before impact. **e**, Back side of perfectly oriented meteorite no. 5 (10.9 g), with a front shell exhibiting thick radially flowing crust and a thinly crusted aft-shell. **f**, The very homogeneous coarse-grained broken surface of large meteorite no. 16 (10 cm diameter).



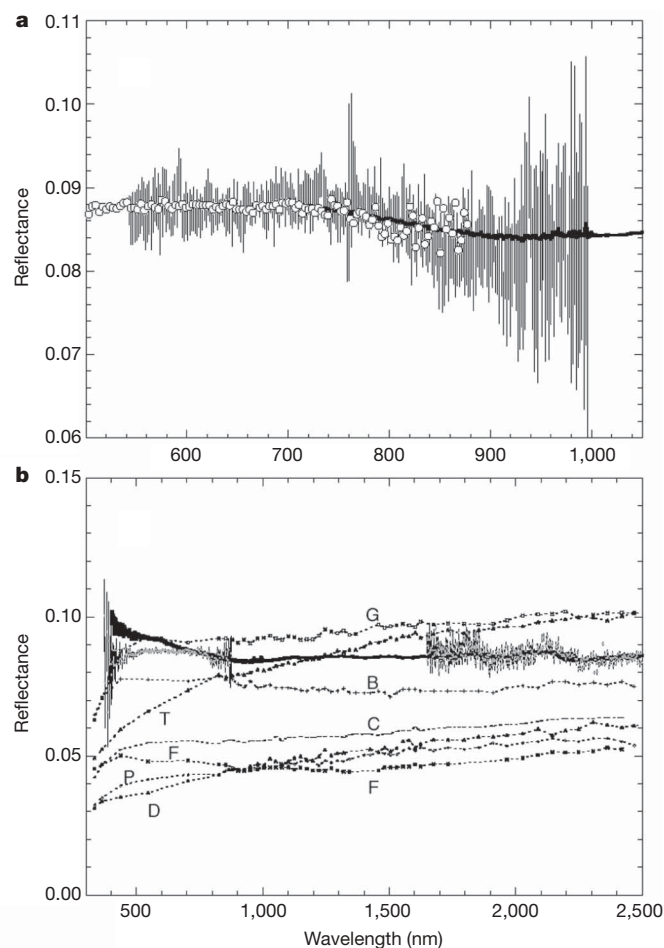


**Figure 3 | Petrography<sup>15</sup> of Almahata Sitta.** **A**, Large-scale back-scattered electron view showing high- and low-porosity lithologies; arrows indicate large carbonaceous inclusions; most olivine and pyroxene aggregates have interstitial silicates whose Si-content increases adjacent to metal grains. Mineral fragments include polycrystalline olivine ( $\text{Fa}_{8-15}$ ;  $\text{CaO} = 0.15\text{--}0.51$  wt%;  $\text{Cr}_2\text{O}_3 = 0.03\text{--}1.58$  wt%), low-calcium pyroxene ( $\text{Fs}_2\text{Wo}_5\text{--Fs}_{17}\text{Wo}_4$ ;  $\text{Cr}_2\text{O}_3 = 0.33\text{--}1.02$  wt%), pigeonite ( $\text{Fs}_{15}\text{Wo}_5\text{--Fs}_{18}\text{Wo}_{11}$ ;  $\text{Cr}_2\text{O}_3 = 0.72\text{--}1.11$  wt%) and carbonaceous aggregates, kamacite ( $\text{Fe}_{0.92}\text{Ni}_{0.08}\text{--Fe}_{0.96}\text{Ni}_{0.04}$ ) and troilite. Some clasts consist of rounded pyroxene grains containing an abundant Fe-rich nanophase. **B**, Low-porosity grains show rounded crystals. Some carbonaceous areas (c) and a few pores (p) are marked. **C**, Pore containing euhedral to anhedral olivine and pyroxene crystals. **D**, A large carbonaceous aggregate containing dispersed, fine-grained troilite and kamacite, the latter containing Si and P. Note the high porosity (p). Raman spectra measure the carbonaceous grains to be amongst the most graphitic of any meteorite yet studied, with a G band centre and full-width at half-maximum of  $1,572 \pm 2.1$  and  $42 \pm 5$   $\text{cm}^{-1}$ , respectively. Imaging Raman shows grain sizes of  $\sim 30$   $\mu\text{m}$  with slightly higher aromatic order near the rim. Two  $10\text{-}\mu\text{m}$ -sized nano-diamonds were imaged in their host material, showing D band peak shifts from latent or biaxial strain. Aliphatic carbon is present too, with weak aliphatic CH-stretch vibration bands peaking at  $2,968$ ,  $2,921$  and  $2,852$   $\text{cm}^{-1}$  (ref. 23).

$83 \pm 25$  t and the kinetic energy of impact  $(6.4 \pm 1.9) \times 10^{12}$  J (at 50 km). This compares well with our estimate calculated from acoustic signals from the fireball detected at the Kenyan infrasonic array I32KE:  $(6.7 \pm 2.1) \times 10^{12}$  J. Analysis of the bolide light curve shows that the total radiated energy was about  $4.0 \times 10^{11}$  J (ref. 7), which translates empirically<sup>19</sup> to a pre-atmospheric kinetic energy of  $\sim 4 \times 10^{12}$  J, in good agreement.

It is unsurprising that such meteorites have not been collected before. The asteroid started to break apart at an altitude of 46–42 km, when the ram pressure was only 0.2–0.3 MPa, and terminated in catastrophic disruption at a pressure of only 1 MPa. The fireball PE-criterion<sup>20</sup>, which uses a fireball's observed end height, velocity, mass and entry angle as a proxy for estimating its physical structure, would make this a IIIb/a-type, normally associated with cometary debris (which tends to disrupt at pressures of  $\leq 0.1$  MPa). In comparison, the unusual Tagish Lake meteorite was similar in initial mass, entry angle, peak luminosity and light-curve shape, but penetrated deeper into the atmosphere, breaking at 40–29 km, with ablation continuing until 27 km (PE = IIIa/II)<sup>21</sup>.

Ureilites were initially thought to derive from S-class asteroids<sup>22</sup> in the Tholen<sup>3</sup> classification of asteroid reflectance spectra. However, the reflectance spectra of 2008 TC<sub>3</sub> and Almahata Sitta meteorite no. 7 are most similar to B or F class asteroids (Fig. 4a). Unlike B-class objects, the meteorite has no hydrated minerals and a modest  $3\text{-}\mu\text{m}$  OH-stretch vibration band. This is indicative of minor adsorbed



**Figure 4 | Meteorite reflectance spectrum compared to that of asteroid 2008 TC<sub>3</sub>.** **a**, The meteorite spectrum (circles and thick black line) is measured at  $3\text{--}7$  nm resolution relative to a diffuse reflectance standard. The asteroid spectrum (shown as vertical lines, representing the s.d. of each set of 10 measured points) is measured at  $4$  nm resolution relative to the solar analogue star 16 Cyg B. We used the  $4.2$  m William Herschel Telescope and ISIS spectrograph on 6 October at 22:22–22:28 UTC. The Sun–asteroid–Earth phase angle was  $18.6^\circ$ . The asteroid spectrum was scaled vertically to match the albedo of the broken surface of meteorite no. 7 (Fig. 2a). Techniques used to measure the meteorite spectrum:  $\lambda < 700$  nm, freshly broken surface no. 7, using a fibre-fed Ocean Optics spectrometer at an illumination angle of  $20^\circ$  and near-perpendicular viewing (circles);  $\lambda = 350\text{--}2,500$  nm, scraped meteorite surface (thick black line), using a FieldSpec ProFR spectrometer from Analytical Spectral Devices, with reflectance values scaled vertically to match visible albedo data. **b**, Same data (2008 TC<sub>3</sub> shown as grey vertical lines: meteorite no. 7 shown as grey circles and as thick black line over  $350\text{--}2,500$  nm wavelength range) compared to the average reflectance spectra of low albedo asteroid taxonomic classes G, B, C, F, T, P and D<sup>24–27</sup>. Note that individual asteroids within a class show a range of albedo. Long-wavelength near-infrared reflectance was independently measured using a Biorad Excalibur Model 3000 Fourier-transform infrared spectrometer (circles).

telluric water, the meteorite spectra showing none of the substructure diagnostic of many phyllosilicates<sup>23</sup>, and implies that 2008 TC<sub>3</sub> was F class. Other low-albedo asteroid types are redder, while B and G classes have a steep drop-off below  $400$  nm, unlike the meteorite (Fig. 4b)<sup>24–27</sup>. The average asteroid F-class spectrum has a slightly more bluish slope (being more reflective in the blue relative to longer wavelengths) below  $700$  nm, similar to that of a scraped meteorite surface (Fig. 4b), and a slightly steeper slope above  $1,500$  nm.

F-class asteroids comprise only  $\sim 1.3\%$  of asteroids. Backward integrations of Monte Carlo clones of the orbit of 2008 TC<sub>3</sub> show that there is an evolutionary pathway, driven by interactions with

Earth, originating from orbits similar to only one other known F-class asteroid: the 2.6-km sized (152679) 1998 KU<sub>2</sub>. Other candidate parent bodies may be identified in the future.

Received 6 February; accepted 20 February 2009.

- Lauretta, D. S. & McSween, H. Y. Jr (eds) *Meteorites and the Early Solar System II* (Univ. Arizona Press, 2006).
- Vernazza, P. et al. Compositional differences between meteorites and near-Earth asteroids. *Nature* **454**, 858–860 (2008).
- Tholen, D. J. in *Asteroids II* (eds Matthews, M. S., Binzel, R. P. & Gehrels, T.) 1139–1150 (Univ. Arizona Press, 1989).
- Kowalski, R. A. et al. in *MPEC 2008-T50* (ed. Williams, G. V.) 1–1 (Minor Planet Center, Smithsonian Astrophysical Observatory, 2008).
- Yeomans, D. NASA/JPL Near-Earth Object Program Office Statement (<http://neo.jpl.nasa.gov/news/news159.html>) (6 October 2008).
- Chesley, S., Chodas, P. & Yeomans, S. NASA/JPL Near-Earth Object Program Office Statement (<http://neo.jpl.nasa.gov/news/2008tc3.html>) (4 November 2008).
- Pravec, P. et al. Tumbling asteroids. *Icarus* **173**, 108–131 (2005).
- Brown, P. G. *US Government release: Bolide detection notification 2008–282* (15 October 2008); (<http://aquarid.physics.uwo.ca/~pbrown/usaf/usg282.txt>).
- Borovicka, J. & Charvat, Z. 2008 TC<sub>3</sub>. *IAU Circ. No.* 8994 (2008).
- Swinbank, R. & O'Neill, A. A stratosphere-troposphere data assimilation system. *Mon. Weath. Rev.* **122**, 686–702 (1994).
- Elhassan, M., Shaddad, M. H. & Jenniskens, P. On the trail of 2008 TC<sub>3</sub>. (Astronomy Picture of the Day, NASA Goddard Space Flight Center, 8 November 2008); (<http://apod.nasa.gov/apod/ap081108.html>).
- Clayton, R. N. & Mayeda, T. K. Oxygen isotope studies of achondrites. *Geochim. Cosmochim. Acta* **60**, 2681–2708 (1996).
- Goodrich, C. A., Van Orman, J. A. & Wilson, L. Fractional melting and smelting on the ureilite parent body. *Geochim. Cosmochim. Acta* **71**, 2876–2895 (2007).
- Mittlefehldt, D. W., McCoy, T. J., Goodrich, C. A. & Kracher, A. Non-chondritic meteorites from asteroidal bodies. *Rev. Mineral.* **36**, 1–195 (1998).
- Zolensky, M. et al. Andreyivanovite: A second new phosphide from the Kaidun meteorite. *Am. Mineral.* **93**, 1295–1299 (2008).
- Goodrich, C. A. Ureilites: A critical review. *Meteoritics* **27**, 327–353 (1992).
- Britt, D. T. & Consolmagno, S. J. Stony meteorite porosities and densities: A review of the data through 2001. *Meteorit. Planet. Sci.* **38**, 1161–1180 (2003).
- Pravec, P. & Harris, A. W. Binary asteroid population. I. Angular momentum content. *Icarus* **190**, 250–259 (2007).
- Brown, P., Spalding, R. E., ReVelle, D. O., Tagliaferri, E. & Worden, S. P. The flux of small near-Earth objects colliding with the Earth. *Nature* **420**, 294–296 (2002).
- Ceplecha, Z. et al. Meteor phenomena and bodies. *Space Sci. Rev.* **84**, 327–471 (1998).
- Brown, P. G., ReVelle, D. O., Tagliaferri, E. & Hildebrand, A. R. An entry model for the Tagish Lake fireball using seismic, satellite and infrasound records. *Meteorit. Planet. Sci.* **37**, 661–675 (2002).
- Gaffey, M. J. et al. Mineralogic variations within the S-type asteroid class. *Icarus* **106**, 573–602 (1993).
- Sandford, S. A. The mid-infrared transmission spectra of Antarctic ureilites. *Meteoritics* **28**, 579–585 (1993).
- Hiroi, T., Zolensky, M. E. & Pieters, C. M. Discovery of the first D-asteroid spectral counterpart: Tagish Lake meteorite. *Lunar Planet. Sci. Conf.* **32**, abstr. 1776 (2001).
- Tholen, D. J. *Asteroid Taxonomy from Cluster Analysis of Photometry*. Ph.D. Thesis, Univ. Arizona (1984).
- Zellner, B., Tholen, D. J. & Tedesco, E. F. The eight-color asteroid survey: Results for 589 minor planets. *Icarus* **61**, 335–416 (1985).
- Bell, J. F. Mineralogical clues to the origins of asteroid dynamical families. *Icarus* **78**, 426–440 (1989).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank the University of Khartoum for support of the field campaigns, and students and staff of the Physics Department of the Faculty of Sciences for their efforts to recover the meteorites. P.J. is supported by the NASA Planetary Astronomy program. D.R. acknowledges the support of NASA's Cosmochemistry program (grant NNX07AI48G). A. Alunni, J. Travis-Garcia and L. Hofland of NASA Ames Research Center, and J. Herrin of NASA Johnson Space Flight Center, provided laboratory assistance. The work conducted at JPL/Caltech was under contract with NASA. The William Herschel Telescope is operated on the island of La Palma by the Isaac Newton Group in the Spanish Observatorio del Roque de los Muchachos of the Instituto de Astrofísica de Canarias.

**Author Contributions** P.J., M.H.S., D.N., S.E. and A.M.K. led the field search for meteorites. M.E.Z., L.L. and G.A.R. performed the petrographic analysis. J.M.F. performed the trace element bulk chemistry analysis. D.R. performed the oxygen isotope analysis. A.S. performed the Raman analysis. S.R.C. calculated the orbit and ground track. A.F., S.D., H.H.H. and G.R. observed and analysed the astronomical spectrum. P.G.B. analysed the infrasound data. P.G.B., W.N.E. and P.J. performed dark-flight calculations. S.P.W., E.T., M.B.B. and R.E.S. facilitated and analysed the US Government satellite observations. R.D. and M.K. observed the asteroid light curve, which was analysed by P.P., J.B. and Z.C. analysed the Meteosat 8 observations. P.J. and J.V. investigated the link with possible other parent bodies. J.K. provided wind model data. J.A. and P.J. analysed train wind drift. J.L.B. and P.J. measured reflection spectra of the meteorite. R.L.M. and P.J. obtained optical imaging of the meteorite. S.A.S., S.N.M., M.N. and P.J. performed the mid-infrared analysis.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to P.J. (Petrus.M.Jenniskens@nasa.gov).



# Electromotive force and huge magnetoresistance in magnetic tunnel junctions

Pham Nam Hai<sup>1</sup>, Shinobu Ohya<sup>1,2</sup>, Masaaki Tanaka<sup>1,2</sup>, Stewart E. Barnes<sup>3,4</sup> & Sadamichi Maekawa<sup>5,6</sup>

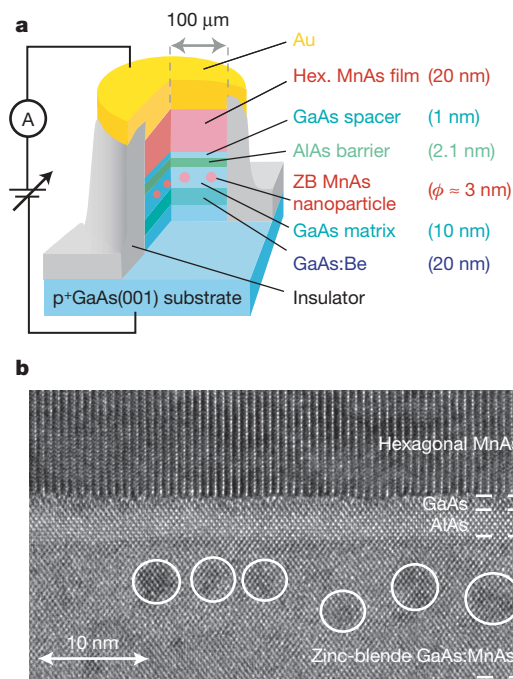
The electromotive force (e.m.f.) predicted by Faraday's law reflects the forces acting on the charge,  $-e$ , of an electron moving through a device or circuit, and is proportional to the time derivative of the magnetic field. This conventional e.m.f. is usually absent for stationary circuits and static magnetic fields. There are also forces that act on the spin of an electron; it has been recently predicted<sup>1,2</sup> that, for circuits that are in part composed of ferromagnetic materials, there arises an e.m.f. of spin origin even for a static magnetic field. This e.m.f. can be attributed to a time-varying magnetization of the host material, such as the motion of magnetic domains in a static magnetic field, and reflects the conversion of magnetic to electrical energy. Here we show that such an e.m.f. can indeed be induced by a static magnetic field in magnetic tunnel junctions containing zinc-blende-structured MnAs quantum nanomagnets. The observed e.m.f. operates on a timescale of approximately  $10^2$ – $10^3$  seconds and results from the conversion of the magnetic energy of the superparamagnetic MnAs nanomagnets into electrical energy when these magnets undergo magnetic quantum tunnelling. As a consequence, a huge magnetoresistance of up to 100,000 per cent is observed for certain bias voltages. Our results strongly support the contention that, in magnetic nanostructures, Faraday's law of induction must be generalized to account for forces of purely spin origin. The huge magnetoresistance and e.m.f. may find potential applications in high sensitivity magnetic sensors, as well as in new active devices such as 'spin batteries'.

Three ingredients are important to the observation of a large spin-derived e.m.f. The first is an ensemble of superparamagnetic nanometre-sized magnets with a large spin  $S \approx 200$ . Owing to a very large magnetic anisotropy, the magnetic moment  $\mu$  is aligned along the  $z$  direction with a component  $S_z = \pm S$  of the spin in this direction. A static magnetic field  $\mathbf{H} = H\mathbf{z}$  splits these two ground states (with  $S_z = \pm S$ ) by an energy  $2\mu H = 2Sg\mu_B H$  (where  $g$  is the  $g$ -factor and  $\mu_B$  is the Bohr magneton). It is this appreciable energy difference that drives the e.m.f. Second, these nanomagnets constitute an essential conductive path through our magnetic tunnel junctions (MTJs), but have such a small capacitance  $C$  that the Coulomb energy  $U = e^2/(2C)$  for adding or removing electrons exceeds the thermal energy  $k_B T$ , effectively blocking sequential electrical conduction<sup>3</sup>. However, as is commonplace, there are spin-flip channels of many-body origin that conduct under this 'Coulomb blockade'. Third, for a temperature  $T = 3$  K, an  $S \approx 200$  nanomagnet would not usually relax within our ten-minute timescale. However, the spin-flip channels mix  $S_z = -S$  with  $-S+1$  and ultimately the two ground states  $S_z = \pm S$ . With the conduction of a single electron, relaxation  $-S \rightarrow S$  occurs, the electron gains an energy  $2Sg\mu_B H$ , and for an ensemble this results in a steady current driven by an e.m.f.  $\mathcal{E} = 2Sg\mu_B H/e$ .

Normally an MTJ consists of metallic thin-film ferromagnetic electrodes and a thin tunnel barrier made of an insulator. The MTJs in this study are unique (Fig. 1a); they consist of zinc-blende-structured

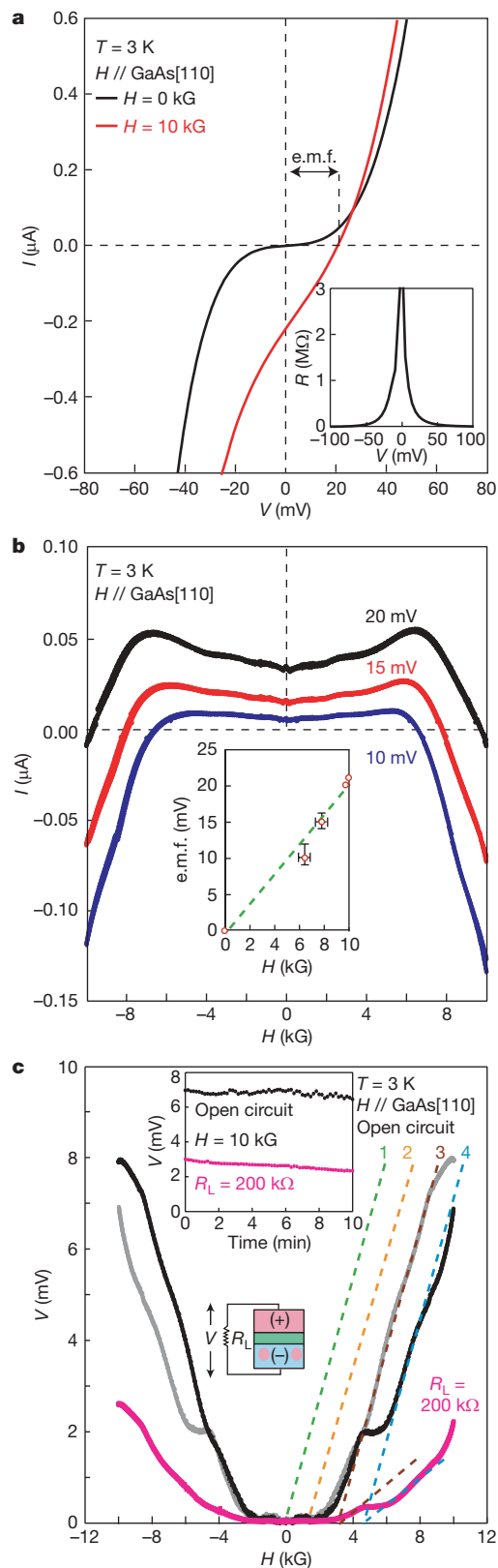
MnAs nanoparticles as a bottom electrode, a 2.1-nm-thick AlAs tunnel barrier with a 1-nm-thick GaAs spacer, and a top electrode comprising a 20-nm-thick hexagonal NiAs-structure MnAs thin film. Zinc-blende MnAs thin films are difficult to fabricate because the bulk phase is not stable<sup>4</sup>. However, using phase separation techniques<sup>5–8</sup>, a GaMnAs alloy ferromagnetic semiconductor layer, when annealed at  $\sim 480^\circ\text{C}$ , generates self-assembled zinc-blende MnAs nanoparticles of diameter 2–3 nm. Figure 1b shows a transmission electron microscopy lattice image of our typical MTJ. Dark areas indicated by white circles correspond to zinc-blende MnAs nanoparticles in which Mn atoms have precipitated in the zinc-blende crystal structure. Spin-dependent tunnelling processes occur between these MnAs nanoparticles and the top NiAs-structured MnAs film through the AlAs tunnel barrier.

Figure 2a shows current–voltage ( $I$ – $V$ ) characteristics of a round, 200- $\mu\text{m}$ -diameter MTJ, measured at 3 K; data are shown for zero



**Figure 1 | Device structure.** **a**, Schematic structure of our magnetic tunnel junctions (MTJs), comprising hexagonal (Hex.) NiAs-structure MnAs (20 nm)/GaAs (1 nm)/AlAs (2.1 nm)/GaAs:MnAs (10 nm) thin films grown on a  $p^+$ GaAs(001) substrate. The GaAs:MnAs film contains zinc-blende-structure (ZB) MnAs nanoparticles.  $\phi$ , diameter. **b**, Transmission electron microscopy lattice image of an MTJ. White circles indicate the areas of zinc-blende MnAs nanoparticles.

<sup>1</sup>Department of Electrical Engineering and Information Systems, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan. <sup>2</sup>Japan Science and Technology Agency, 4-1-8 Honcho, Kawaguchi-shi 332-0012, Japan. <sup>3</sup>Physics Department, University of Miami, Coral Gables, Florida 33124, USA. <sup>4</sup>TCM, Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, UK. <sup>5</sup>Institute for Materials Research, Tohoku University, Sendai 980-8577, Japan. <sup>6</sup>CREST, Japan Science and Technology Agency, Tokyo 100-0075, Japan.



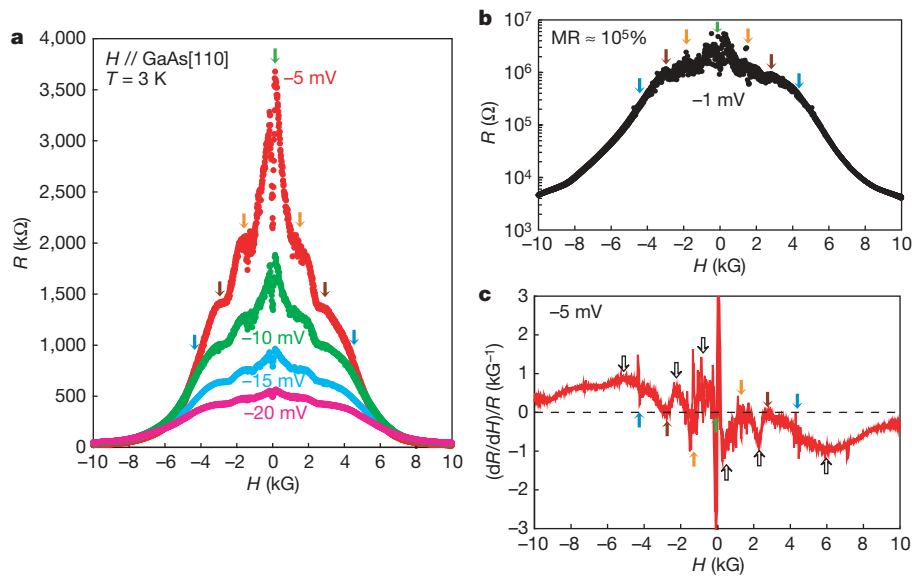
magnetic field (black curve), and for a 10 kG magnetic field (red curve) applied along the easy axis of magnetization of the top MnAs thin film (that is, along the  $[1\bar{1}20]$  MnAs and  $[110]$  GaAs axes in the film plane). The polarity of the applied bias voltage is defined as the voltage of the top MnAs film with respect to the substrate. The  $I$ - $V$  curve, for 10 kG, is shifted towards the positive bias voltage region by 21 mV. This shift corresponds to an e.m.f. of 21 mV generated by the

**Figure 2 | Transport characteristics of an MTJ.** **a**, Current-voltage ( $I$ - $V$ ) characteristics of a 200- $\mu\text{m}$ -diameter MTJ, measured at 3 K without (black curve) and with (red curve) a magnetic field of 10 kG applied in plane along the GaAs $[110]$  azimuth. In a magnetic field of 10 kG, the origin of the  $I$ - $V$  curve was shifted towards positive bias by 21 mV, corresponding to an e.m.f. generated by the MTJ. Inset, voltage dependence of the MTJ resistance at zero field. A Coulomb blockade of  $U \approx 50$  meV was observed. **b**, Current versus magnetic field of an MTJ measured at 10, 15 and 20 mV. Inset, induced e.m.f. as a function of magnetic field  $H$  measured by the compensation method (see text). Error bars reflect a systematic uncertainty due to the time dependence of the e.m.f. **c**, Voltage across an external resistance ( $R_L$ ) connected directly to the MTJ in a magnetic field without a current source. Black and red curves are the output voltages of the MTJ when the external resistance is respectively infinitely large (that is, open circuit) and 200 k $\Omega$ . The grey curve is the black curve but reflected about the vertical axis. The inset shows that the voltage lasts for at least 10 min.

MTJ. With a static magnetic field, this e.m.f. causes the top MnAs film and the bottom MnAs nanoparticles to have positive and negative potentials, respectively. Effects of the e.m.f., induced by a static magnetic field, can also be clearly seen in Fig. 2b, which shows the magneto-current measured at positive bias voltages smaller than 21 mV. For low fields, the current is positive and increases with increasing magnetic field, revealing an ordinary negative magnetoresistance. However, for fields greater than about 5 kG, the current decreases with increasing magnetic field, and eventually becomes negative. The reversed current at 10 kG indicates that the induced e.m.f. is larger than the applied bias voltage. When the external voltage compensates the internal e.m.f., no current flows in the circuit. The inset in Fig. 2b shows the magnetic-field dependence of the induced e.m.f. measured by this compensation method.

Figure 2c shows the voltage drop across an external resistance ( $R_L$ ) connected directly to the MTJ as a function of magnetic field ('spin battery' configuration). The black curve is the output voltage of the MTJ when the external resistance is infinitely large (that is, open circuit), and the red curve is the voltage drop across an external resistance of 200 k $\Omega$ . For finite fields, a voltage appears across the load resistance, demonstrating that the MTJ indeed constitutes a 'spin battery'. As seen in the inset of Fig. 2c, the voltage lasted for at least 10 min, implying a decay constant  $\tau \approx 10^2$ - $10^3$  s. When the magnetic field was turned off, the voltage immediately disappeared, showing that the intrinsic time constant  $RC$  of the MTJ is short. Indeed, using a resistance  $R \approx 10^6 \Omega$  and a capacitance  $C \approx NC_0$ —where the number of nanoparticles is  $N \approx 10^9$  and the capacitance of a single nanoparticle is  $C_0 = e^2/2U \approx 10^{-18}$  F ( $U = 50$  meV, as discussed later, is the Coulomb energy of zinc-blende MnAs nanoparticles)— $RC$  is expected to be about  $10^{-3}$  s. The induced magnetic e.m.f. continually pumps charges through the nanoparticles, thereby maintaining the output voltage. Thermoelectric effects are not responsible for the observed e.m.f., as discussed in Supplementary Information.

Figures 3a and 3b show the magnetoresistance of the MTJ measured with bias voltages  $V$  of  $-20$ ,  $-15$ ,  $-10$ ,  $-5$  and  $-1$  mV. For  $V = -20$  mV, a large magnetoresistance ratio (defined as  $(R_{\text{max}} - R_{10\text{kG}})/R_{10\text{kG}}$ ) of 1,220% was observed. This ratio is equal to  $(I_{10\text{kG}} - I_{\text{min}})/I_{\text{min}}$ , which is the appropriate figure of merit for this device when used as a magnetic field sensor. With decreasing  $V$ , the magnetoresistance ratio increased, reaching 100,000% for the smallest bias voltage ( $-1$  mV). This is three orders of magnitude greater than conventional giant magnetoresistance (GMR)<sup>9,10</sup> or tunnelling magnetoresistance (TMR)<sup>11-14</sup>, and comparable to the colossal magnetoresistance (CMR) observed in some manganese perovskites<sup>15,16</sup>. An even larger magnetoresistance ratio might be obtained if the junction was biased with a voltage smaller than  $-1$  mV. This huge magnetoresistance is a consequence of the combined effect of the induced e.m.f. and Coulomb blockade in our zinc-blende MnAs nanoparticles. The inset in Fig. 2a shows that the resistance of the MTJ increased sharply when  $|V| < 50$  mV, implying a Coulomb energy  $U \approx 50$  meV. The resistance at small bias is therefore quite large, as electron tunnelling through the



**Figure 3 | Huge magnetoresistance.** Magnetoresistance of the MTJ measured at  $V = -20, -15, -10, -5$  mV (**a**) and  $-1$  mV (**b**). A huge magnetoresistance (MR) as large as 100,000% is observed. This huge value is a consequence of the induced e.m.f. combined with the Coulomb blockade

nanoparticles is almost suppressed by the Coulomb blockade. The e.m.f. induced by a magnetic field causes the MTJ to be self-biased in the same direction as the applied bias voltage, with a concomitant dramatic reduction of the effective resistance.

An e.m.f. induced by a static magnetic field was first theoretically predicted for magnetic nanostructures, such as magnetic nanowires or spin-valves<sup>1,2</sup>. Viewed macroscopically, this motive force reflects the energy conservation requirements<sup>17,18</sup>. A domain of magnetic moment  $\mu$  has an energy  $+\mu H$  when anti-parallel to the field but  $-\mu H$  when parallel. In an anti-parallel field, an electron induces the transition  $+\mu \rightarrow -\mu$ . The magnetic energy change  $2\mu H$  is transferred to the electrical sector, generating an e.m.f. which drives a current in an external load<sup>17</sup>.

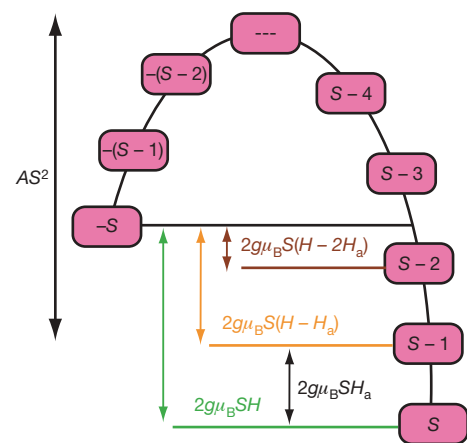
The experimental Coulomb blockade energy of 50 meV is sufficiently large that it almost totally suppresses transport via the single-electron states. The result is a very large zero field resistance. In addition to these electronic levels, there is a set of levels reflecting the spin state  $|S_z\rangle$ , where  $S_z$  is the angular momentum quantum number of the nanomagnet order parameter of magnitude  $S$ . Such nanomagnets have been much studied<sup>19</sup>. Figure 4 illustrates the relevant energy scheme for a dominant uniaxial anisotropy, and the well-known phenomenology is briefly described in the legend. Electrons can tunnel through the nanomagnet to the bottom electrode, despite the blockade, provided that the magnetization ( $S$  state) of the nanomagnet also tunnels between the magnetic levels of Fig. 4. If the electron is added to an initial magnetic state  $| -S \rangle$ , this can couple to an entirely similar but empty state with the equivalent magnetic ground state  $| S \rangle$  and which differs in energy only by the relatively small Zeeman term  $2g\mu_B SH$ . The theory, outlined in Supplementary Information, shows that this amount of energy is given to, or taken from, the tunnelling electron, resulting in an e.m.f.

Equating our open circuit e.m.f. of 21 mV with  $2g\mu_B SH/e$  implies that  $S \approx 200$ . With increasing field  $H$ , when the energy of the state  $| -S \rangle$  is greater than that of  $| S-1 \rangle$ , it becomes possible to conduct using these states, which involves tunnelling from  $-S$  to  $S-1$ . Now the magnetic matrix elements are larger, and the conductivity greater. The e.m.f. now corresponds to the difference in energy of these states, as in Fig. 4, and is reduced by  $2g\mu_B SH_a$ , where  $H_a$  is the anisotropy field. Depending upon the real or effective load, with increasing  $H$  many new potential e.m.f. branches open as a function of  $H$ , and each is displaced relative to its neighbour by  $2g\mu_B SH_a$ . Clearly, changes in

effect. **c**,  $(dR/dH)/R$  at  $V = -5$  mV. There are evident peaks corresponding to the level crossings (see text) around 0, 1.5, 3.0 and 4.5 kG, marked by the coloured arrows. The white arrows show the positions of steep change of resistance.

the number of channels and in mobility, which occur at each level crossing, are reflected in transport properties. When the field direction is reversed, the dominant tunnelling direction of magnetization is also reversed. Therefore, an initial asymmetric population of zinc-blende MnAs nanomagnets, with opposite magnetization directions, will result in an asymmetric  $V(H)$  curves (Fig. 2c).

We examine the e.m.f. data for branches with identical slopes,  $2g\mu_B S$ , but displaced by a constant amount  $2g\mu_B SH_a$ , and we also look for sudden changes in resistance with this same period. In Fig. 3a we indeed see there are such changes in the resistance and that they are separated by approximately 1.5 kG (corresponding to  $H_a$ ). In Fig. 3c we show the resistance derivative with evident equally spaced peaks indicated by the coloured arrows. The e.m.f. data (Fig. 2c) show equally evident branches. For positive  $H$  there are two e.m.f. open circuit (black curve) branches, which extrapolate



**Figure 4 | Magnetic energy.** The magnetic levels have, approximately, an energy  $-AS_z^2$ , with  $A$  being the uniaxial magnetic anisotropy constant. There are two equivalent ground states  $|\pm S\rangle$  separated by a large energy barrier  $\sim AS^2$ . Magnetic tunnelling occurs between the states  $|\pm S\rangle$ . A magnetic field adds  $g\mu_B S_z H$  to the energy and lifts the state  $| -S \rangle$  above  $| S \rangle$  by  $2g\mu_B SH$  and effectively suppresses tunnelling. Level crossings, and tunnelling, occur for periodic values of field with period  $2g\mu_B SH_a$ , where  $H_a = A/g\mu_B$  is the anisotropy field. At such crossings the magnetic relaxation rate is enhanced<sup>19</sup>. In our experiments these same level crossings open new transport channels and lead to new e.m.f. branches.



back to approximately 3.0 kG and 4.5 kG. Only the 3.0 kG branch is evident for negative field. Because the external load is then zero, the first missing branch appears when the internal e.m.f. is compensated by an external bias. The data are shown in the inset of Fig. 2b and roughly agree with the extrapolation of the green dashed line (labelled 1) in Fig. 2c. When a load of 200 k $\Omega$  is added, the e.m.f. is reduced, but it remains the case that the similarly identified branches extrapolate back to the same fields, which in turn correspond to the coloured arrows in the transport data shown in Fig. 3; this lends strong support to our model of magnetic co-tunnelling. The relaxation times of superparamagnetic systems are notoriously long<sup>19</sup> and quite compatible with the observed lifetime ( $\sim 10^2$ – $10^3$  s) of our e.m.f. The magnetization measurement results described in Supplementary Information strongly support our interpretation of the behaviour of zinc-blende MnAs nanomagnets.

We have shown that a static magnetic field results in an e.m.f. compatible with predictions. This cannot reflect a time derivative of the magnetization ( $\partial M/\partial t$ ) and the usual Faraday's law of induction as this would change sign with the applied field. It lends strong support to the contention that, in magnetic nanostructures, Faraday's law must reflect forces due to both the electric and magnetic dipole charges of the electrons. The resulting huge magnetoresistance together with the e.m.f. may have potential applications as a completely new kind of magnetic sensor with ultrahigh sensitivity, and as active devices such as 'spin batteries'.

## METHODS SUMMARY

Our MTJs consist of ferromagnetic hexagonal NiAs-structure MnAs (20 nm)/GaAs (1 nm)/AlAs (2.1 nm)/GaAs:MnAs (10 nm) thin films, grown by molecular beam epitaxy (MBE) on p<sup>+</sup>GaAs(001) substrates. To fabricate the zinc-blende MnAs nanoparticles, we grew a 10-nm-thick Ga<sub>0.94</sub>Mn<sub>0.06</sub>As thin film at 240 °C and then annealed at 480 °C for 20 min in the MBE growth chamber, during which phase separation occurred in the GaMnAs layer and the zinc-blende MnAs nanoparticles of diameter 2–3 nm were formed in the GaAs matrix. After the growth, we fabricated round, 200- $\mu$ m-diameter mesa diode structures for transport measurements. The MTJs were mounted and cooled to 3 K in a refrigerator unit (JANIS Research) equipped with an external electromagnet (0–10 kG). Transport measurements were carried out with a conventional two-terminal direct-current (d.c.) method using an Advantest R6246 two-channel voltage/current source/monitor. An Advantest R6871E multi-meter was used for monitoring the voltage drop across the external resistance connected to the MTJ with a magnetic field (the 'spin battery' configuration). The sweep rate of the magnetic field was 10 kG min<sup>-1</sup>.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 5 August 2008; accepted 12 February 2009.

Published online 8 March 2009.

1. Barnes, S. E. & Maekawa, S. Generalization of Faraday's law to include nonconservative spin forces. *Phys. Rev. Lett.* **98**, 246601 (2007).
2. Barnes, S. E. Spin motive forces, "measurements", and spin-valves. *J. Magn. Magn. Mater.* **310**, 2035–2037 (2007).

3. Fulton, T. A. & Dolan, G. J. Observation of single-electron charging effects in small tunnel junctions. *Phys. Rev. Lett.* **59**, 109–112 (1987).
4. Sanvito, S. & Hill, A. N. Ground state of half-metallic zinc-blende MnAs. *Phys. Rev. B* **62**, 15553–15560 (2000).
5. Sato, K., Katayama-Yoshida, H. & Dederichs, P. H. High Curie temperature and nano-scale spinodal decomposition phase in dilute magnetic semiconductors. *Jpn. J. Appl. Phys.* **44**, L948–L951 (2005).
6. Moreno, M., Trampert, A., Jenichen, B., Daweritz, L. & Ploog, K. H. Correlation of structure and magnetism in GaAs with embedded Mn(Ga)As magnetic nanoclusters. *J. Appl. Phys.* **92**, 4672–4677 (2002).
7. Yokoyama, M., Yamaguchi, H., Ogawa, T. & Tanaka, M. Zinc-blende-type MnAs nanoclusters embedded in GaAs. *J. Appl. Phys.* **97**, 10D317 (2005).
8. Kwiatkowski, A. et al. Structure and magnetism of MnAs nanocrystals embedded in GaAs as a function of post-growth annealing temperature. *J. Appl. Phys.* **101**, 113912 (2007).
9. Binasch, G., Grünberg, P., Saurenbach, F. & Zinn, W. Enhanced magnetoresistance in layered magnetic structures with antiferromagnetic interlayer exchange. *Phys. Rev. B* **39**, 4828–4830 (1989).
10. Baibich, M. N. et al. Giant magnetoresistance of (001)Fe/(001)Cr magnetic superlattices. *Phys. Rev. Lett.* **61**, 2472–2475 (1988).
11. Moodera, J. S., Kinder, L. R., Wong, T. M. & Meservey, R. Large magnetoresistance at room temperature in ferromagnetic thin film tunnel junctions. *Phys. Rev. Lett.* **74**, 3273–3276 (1995).
12. Miyazaki, T. & Tezuka, N. Giant magnetic tunnelling effect in Fe/Al<sub>2</sub>O<sub>3</sub>/Fe junction. *J. Magn. Magn. Mater.* **139**, L231–L234 (1995).
13. Parkin, S. S. P. Giant tunnelling magnetoresistance at room temperature with MgO (100) tunnel barriers. *Nature Mater.* **3**, 862–867 (2004).
14. Yuasa, S., Nagahama, T., Fukushima, A., Suzuki, Y. & Ando, K. Giant room-temperature magnetoresistance in single-crystal Fe/MgO/Fe magnetic tunnel junctions. *Nature Mater.* **3**, 868–871 (2004).
15. Jin, S. et al. Thousandfold change in resistivity in magnetoresistive La-Ca-Mn-O films. *Science* **264**, 413–415 (1994).
16. Tokura, Y. (ed.) *Advances in Condensed Matter Science Vol. 2, Colossal Magnetoresistance Oxides* (Gordon & Breach, 2000).
17. Barnes, S. E., Ieda, J. & Maekawa, S. Magnetic memory and current amplification devices using moving domain walls. *Appl. Phys. Lett.* **89**, 122507 (2006).
18. Barnes, S. E. & Maekawa, S. Currents induced by domain wall motion in thin ferromagnetic wires. Preprint at (<http://arxiv.org/abs/cond-mat/0410021v1>) (2004).
19. Barbara, B. et al. Mesoscopic quantum tunnelling of magnetization. *J. Magn. Magn. Mater.* **140–144**, 1825–1828 (2002).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was partly supported by Grant-in-Aids for Scientific Research No. 18106007, No. 19048018 and No. 20686002, the Special Coordination Programs for Promoting Science and Technology, and R&D for Next-Generation Information Technology by MEXT, PRESTO of JST, and EPSRC (UK). We thank B.-H. Yu for his help in the transport measurements. P.N.H. acknowledges a JSPS Research Fellowship for Young Scientists and the Global COE Program (CO4).

**Author Contributions** P.N.H. designed the experiment, fabricated the samples, collected most of data and performed analysis of data; S.O. set up measurement apparatuses and gave experimental advice; M.T. managed and planned the research and supervised the experiment; and S.E.B. and S.M. developed the theoretical explanation of the experiment. All authors discussed the results and commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to M.T. ([masaaki@ee.t.u-tokyo.ac.jp](mailto:masaaki@ee.t.u-tokyo.ac.jp)).

## METHODS

Our MTJs consist of ferromagnetic hexagonal NiAs-structure MnAs (20 nm)/GaAs (1 nm)/AlAs (2.1 nm)/GaAs:MnAs (10 nm) thin films. The GaAs:MnAs film contains zinc-blende-structure MnAs nanoparticles in a GaAs matrix (Fig. 1a). These structures are grown by MBE on  $p^+$ -GaAs(001) substrates. First, we grew a 20-nm-thick Be-doped GaAs buffer layer on a  $p^+$ -GaAs(001) substrate at 580 °C. After cooling the substrate temperature to 240 °C, we grew a 10-nm-thick  $\text{Ga}_{0.94}\text{Mn}_{0.06}\text{As}$  film, a 2.1-nm-thick AlAs tunnel barrier and then a 1-nm-thick GaAs spacer. Then the structure was annealed at 480 °C for 20 min in the MBE growth chamber, during which phase separation occurred in the GaMnAs layer and the zinc-blende MnAs nanoparticles of diameter 2–3 nm were formed in the GaAs matrix. When annealed at higher temperature ( $>550$  °C), the more stable hexagonal NiAs phase precipitates. By inserting a very thin GaAs layer between AlAs and MnAs, the quality of the top MnAs film is improved. Furthermore, by observing the reconstruction in the reflection high-energy electron diffraction (RHEED) patterns of the GaAs layer, we can control the annealing temperature more precisely. Finally, a 20-nm-thick type-A hexagonal NiAs-structure MnAs thin film was grown at 240 °C as a top electrode. Here, the epitaxial relationship of type-A MnAs on the semiconductor layer is  $(\bar{1}100)\text{MnAs} // (001)\text{GaAs}$ ,  $[\bar{1}\bar{1}20]\text{MnAs} // [110]\text{GaAs}$ , and  $[0001]\text{MnAs} // [110]\text{GaAs}$ , and the easy magnetization axis is  $[\bar{1}\bar{1}20]\text{MnAs}$  in the film plane<sup>20,21</sup>. After completing the growth, post-growth annealing was carried out in the growth chamber at 310 °C for 10 min to improve the crystal quality of the top MnAs film.

After the initial growth, we fabricated, by standard photolithography and chemical etching, round, 200- $\mu\text{m}$ -diameter mesa diode structures. We spin-coated an insulating negative resist on the sample, opened a contact hole with 180  $\mu\text{m}$  diameter on top of the mesa, and fabricated a metal electrode by evaporating Au on this surface. The MTJs were mounted and cooled to 3 K in a refrigerator unit (JANIS Research) equipped with an external electromagnet (0–10 kG). Transport measurements were carried out with a conventional two-terminal direct-current (d.c.) method using an Advantest R6246 two-channel voltage/current source/monitor. An Advantest R6871E multi-meter was used for monitoring the voltage drop across the external resistance connected to the MTJ with a magnetic field (the ‘spin battery’ configuration). The sweep rate of the magnetic field was 10 kG  $\text{min}^{-1}$ . The results presented here (e.m.f. and huge magnetoresistance) were observed in many MTJs fabricated by the same process.

20. Tanaka, M. *et al.* Epitaxial orientation and magnetic properties of MnAs thin films grown on (001) GaAs: Template effects. *Appl. Phys. Lett.* **65**, 1964–1966 (1994).
21. Tanaka, M. Ferromagnet (MnAs)/III–V semiconductor hybrid structures. *Semicond. Sci. Technol.* **17**, 327–341 (2002).

# Variable Quaternary chemical weathering fluxes and imbalances in marine geochemical budgets

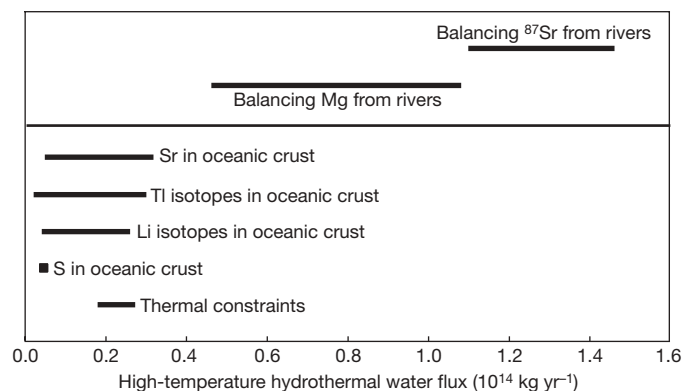
Derek Vance<sup>1</sup>, Damon A. H. Teagle<sup>2</sup> & Gavin L. Foster<sup>1</sup>

Rivers are the dominant source of many elements and isotopes to the ocean. But this input from the continents is not balanced by the loss of the elements and isotopes through hydrothermal and sedimentary exchange with the oceanic crust, or by temporal changes in the marine inventory for elements that are demonstrably not in steady state<sup>1–4</sup>. To resolve the problem of the observed imbalance in marine geochemical budgets, attention has been focused on uncertainties in the hydrothermal and sedimentary fluxes<sup>1–4</sup>. In recent Earth history, temporally dynamic chemical weathering fluxes from the continents are an inevitable consequence of periodic glaciations<sup>5–9</sup>. Chemical weathering rates on modern Earth are likely to remain far from equilibrium owing to the physical production of finely ground material at glacial terminations<sup>10–13</sup> that acts as a fertile substrate for chemical weathering. Here we explore the implications of temporal changes in the riverine chemical weathering flux for oceanic geochemical budgets. We contend that the riverine flux obtained from observations of modern rivers is broadly accurate, but not representative of time-scales appropriate for elements with oceanic residence longer than Quaternary glacial–interglacial cycles. We suggest that the pulse of rapid chemical weathering initiated at the last deglaciation has not yet decayed away and that weathering rates remain about two to three times the average for an entire late Quaternary glacial cycle. Taking into account the effect of the suggested non-steady-state process on the silicate weathering flux helps to reconcile the modelled marine strontium isotope budget with available data. Overall, we conclude that consideration of the temporal variability in riverine fluxes largely ameliorates long-standing problems with chemical and isotopic mass balances in the ocean.

For many elements and isotopes, the riverine flux is the dominant input to the oceans and is at least partially balanced by outputs to—or exchange with—the oceanic crust, either via alteration of basalts or deposition in sediments<sup>1–4</sup>. For some elements with isotopic systems, such as strontium (Sr) and osmium (Os), the oceans are demonstrably not in perfect steady state<sup>14,15</sup>. For example, for the past few million years the oceanic <sup>87</sup>Sr/<sup>86</sup>Sr ratio has been rising at a rate of 0.00054 Myr<sup>−1</sup>. Therefore, the input of relatively radiogenic Sr from the continents cannot be perfectly balanced by unradiogenic Sr derived from hydrothermal exchange with the oceanic crust<sup>3,16</sup>, which for many elements is the second greatest influence on seawater chemistry after rivers<sup>1</sup>. The extent of Sr isotopic alteration of oceanic crust in drill cores and ophiolites<sup>3</sup> implies a high-temperature hydrothermal fluid flux of around  $(0.5–3) \times 10^{13}$  kg yr<sup>−1</sup> (Fig. 1). This estimate agrees well with those obtained from a range of other elements in altered oceanic crust and overlaps with those from thermal calculations (Fig. 1). But, when combined with the modern measured riverine Sr flux (Supplementary Table 1), such a hydrothermal flux would suggest that seawater <sup>87</sup>Sr/<sup>86</sup>Sr should be evolving at a rate of

0.000425 Myr<sup>−1</sup>, eight times faster than implied by the record of Sr isotopes in sea water<sup>14</sup>. To keep the Sr isotope ratio of sea water evolving at the slow rate which the records suggest, the flux of water through high-temperature hydrothermal systems would need to be  $>10^{14}$  kg yr<sup>−1</sup> (Supplementary Table 2), in marked contrast to the consensus from observations of the alteration state of the oceanic crust for a range of elements and isotope systems (Fig. 1). This problem with marine geochemical budgets is by no means unique to Sr. For Mg, another element whose marine geochemistry is the basis of a prominent palaeoceanographic tool, rivers are also the main input and hydrothermal systems are an important sink<sup>4,17</sup>. The Mg budget of the oceans is also demonstrably out of steady state<sup>4,17</sup>, but even the inclusion of a secular increase in oceanic Mg concentrations through time cannot balance the budget.

To resolve these issues, attention has traditionally focused on uncertainties in hydrothermal fluxes. After all, the riverine input flux is directly and comprehensively observable. Pollution of modern



**Figure 1 | Estimates of the hydrothermal water flux at mid-ocean ridges.** Shown is a comparison of estimates of the high-temperature hydrothermal water flux obtained from a range of observations of the oceanic crust (bottom panel) with those required by the need to balance the riverine input of Mg and radiogenic Sr from the continents (top panel). The range of hydrothermal water fluxes required to balance radiogenic Sr in rivers is derived from calculations in Supplementary Tables 2 and 3. These are a factor of 4–20 higher than is permitted by constraints deriving from the alteration state of the oceanic crust for Sr isotopes<sup>3,28</sup>. Analogous constraints from Mg are looser, and the range for the requirement imposed by the riverine input reflects uncertainty over the extent of steady state in the ocean for Mg and debate over the size of outputs from the oceans other than high-temperature hydrothermal systems<sup>4,17</sup>. Overall, the geochemical (Li and Ti isotopes, elemental S) and geophysical (data compiled in refs 28, 29) constraints on the size of the high-temperature hydrothermal flux from the oceanic crust are rather concordant, at  $(0.1–0.2) \times 10^{13}$  kg yr<sup>−1</sup>, and stand in marked contrast to the factor  $\sim 10$  higher value required by the riverine flux of Sr to the oceans.

<sup>1</sup>Bristol Isotope Group, Department of Earth Sciences, University of Bristol, Wills Memorial Building, Bristol BS8 1RJ, UK. <sup>2</sup>School of Ocean and Earth Science, National Oceanography Centre, University of Southampton, Southampton SO14 3ZH, UK.



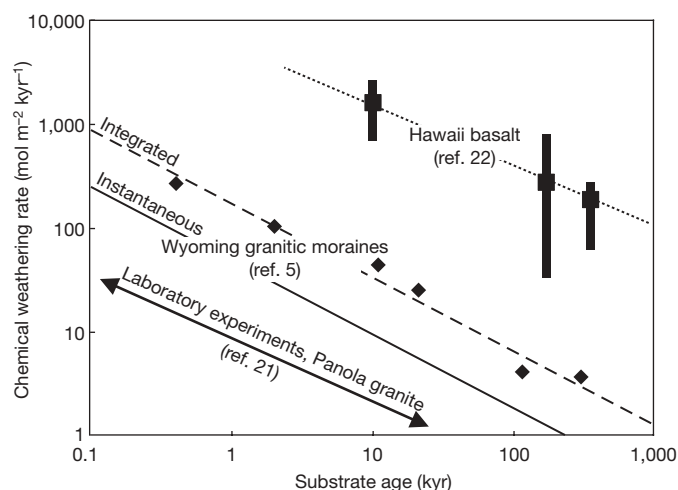
ivers is significant for many elements but not for the main cations at the magnitude of interest here<sup>18</sup>. A potentially more serious issue is the inherent temporal variability of rivers<sup>2,16</sup>, and the possibility that the extensive database for Sr and its isotopes in modern rivers does not accurately yield a global Sr flux to the oceans. However, although major rivers show pronounced seasonality, recent data suggest that the required hydrothermal flux is sensitive only at the 10–15% level to plausible variations in the global riverine flux (Supplementary Tables 1 and 2). The initial conclusion, that the modern riverine Sr flux requires an unrealistically high hydrothermal flux ( $>10^{14}$  kg yr<sup>-1</sup>), around an order of magnitude higher than independent estimates, therefore appears robust. We contend that the riverine flux obtained from observations of modern rivers is broadly accurate, but that it is not representative of timescales of  $10^5$ – $10^7$  years, those appropriate for an element with an oceanic residence time much longer than Quaternary glacial–interglacial cycles.

About 25–30% of the currently ice-free continental area was covered by large ice sheets during the Quaternary glacial periods, including the Last Glacial Maximum 18 kyr ago<sup>19</sup>. Chemical weathering rates underneath the dry-based late Quaternary ice sheets were near negligible<sup>20</sup>. In contrast, glacial and ice sheet physical weathering rates are very high<sup>10,20</sup>; they are estimated to be 10 times greater than without Northern Hemisphere glaciation<sup>10</sup>. The demise of the continental ice sheets at glacial terminations leaves behind large amounts of fresh, finely ground rock in young moraines and soils that provide fertile substrates for chemical weathering. High deglacial physical weathering rates are not confined to high latitudes. In the European Alps, a 12–20-fold increase in sediment delivery to peri-Alpine valleys and lakes<sup>13</sup> occurred 17–11 kyr ago, and is related to dispersal of glacially-ground sediment due to increased melt water and wetter conditions immediately post-glacial. Huge amounts of sediment were delivered to the Ganges floodplain from the Himalaya between 11 and 7 kyr ago<sup>12</sup>, with sediment storage increasing 10 times. This latter phenomenon was driven by the early Holocene peak in monsoonal precipitation<sup>12</sup>, rather than changes in glacial coverage, and there is evidence for a global expression of this phenomenon in the period after about 10 kyr ago<sup>11</sup>.

Field and laboratory observations suggest that newly produced fine-grained material is highly reactive when initially exposed to agents of chemical weathering, but that rates decrease dramatically with time (Fig. 2). The slope of the power law relationship between weathering rate and exposure time for post-glacial soils<sup>5</sup> is similar to that obtained from laboratory experiments on Panola granite<sup>21</sup> and to recent field studies of Hawaiian basalt flows<sup>22</sup> (Fig. 2). It is important to note that for both granitic and basaltic lithologies, and for climatic regimes ranging from cold to tropical, the dependence of weathering rate on substrate age is the same, despite variations in absolute weathering rate. These experiments and observations refer only to silicates. In principle, similar processes could occur with carbonate weathering, but the orders of magnitude faster reaction kinetics for calcite dissolution relative to silicates<sup>20</sup> suggest that carbonate dissolution is much closer to keeping pace with physical grinding. Here we follow others<sup>19</sup> in assuming that carbonate weathering rates remain more or less constant (except on carbonate shelves exposed during sea-level low-stands: see below), and keep pace with environmental change, through glacial–interglacial cycles.

Using the relationship between soil age and silicate weathering rate (Fig. 2), we can estimate the effect of increased high-latitude chemical weathering over the last glacial cycle (Fig. 3a). This indicates that the pulse of rapid chemical weathering initiated at the last deglaciation (18 kyr ago) has not yet decayed away and rates remain about twice the average for an entire late Quaternary glacial cycle. This is supported by measured weathering rates in the drainage basins of two major Canadian rivers, the Mackenzie and the St Lawrence<sup>23</sup>, whose dissolved load suggests a modern weathering rate that is 3–4 times higher than the long-term model rate.

These high modern, relative to long-term, weathering rates are also supported by a range of studies on the modern chemistry of major



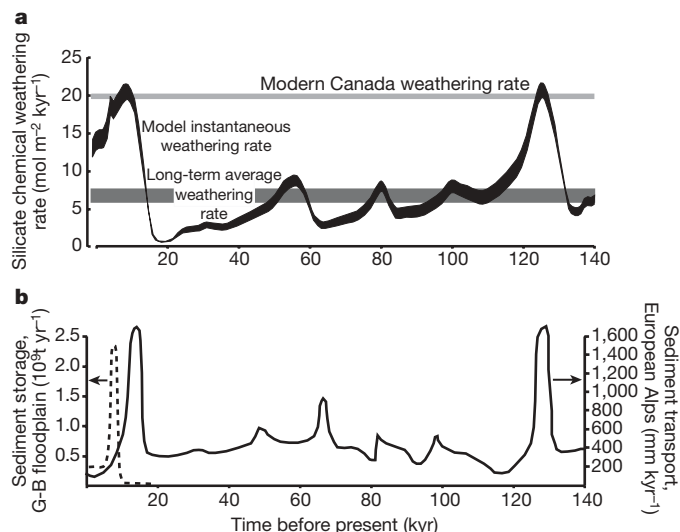
**Figure 2 | Dependence of chemical weathering on substrate age.**

Diamonds, measured chemical depletion rates for a sequence of soils developed on moraines of varying age deposited by retreating glaciers/ice-sheets in Wyoming<sup>5</sup>. The data conform to a power law relationship, with instantaneous weathering rate at any time  $t$  subsequent to exposure from beneath the ice being given by the solid line ( $50t^{-0.71}$ ). Laboratory data on the weathering rate of Panola granite<sup>21</sup> (double-headed arrow, absolute rates depend on starting experimental surface area, exponent =  $-0.61$ ), as well as recent data for a series of Hawaiian basaltic lavas<sup>22</sup> (averages and ranges across a range of precipitation conditions given by squares and vertical bars, exponent =  $-0.61$ ) yield relationships between time-integrated weathering rate and substrate age that are very similar to the Wyoming data. The similarity in the slopes of these relationships suggests a dependence of weathering rate on substrate age that is common across climatic regimes and silicate lithologies, despite a two orders of magnitude variation in absolute weathering rates that is dependent on climate and lithology.

drainage basins, particularly U–Th isotope investigations that often point to departures from steady state in the past 5–20 kyr (data compiled in ref. 9). Moreover, oceanic Pb isotopes have recently provided further support for substantial glacial–interglacial changes in silicate weathering on the continents<sup>9</sup>. Other evidence substantiates suggestions of non-steady-state chemical weathering in some drainage basins not directly affected by glaciation<sup>9</sup>. U–Th isotopes in rivers draining the Andes have, for example, been out of steady state for between 4 and 20 kyr (ref. 9). The exact magnitude of continued departure from steady state in these mountainous regions is important because of their importance to total riverine fluxes<sup>24</sup>, but is much more difficult to quantify than for high latitudes. Furthermore, it is also clear that some lowland tropical rivers are close to steady state at the present day<sup>9</sup>. Nonetheless, the similarity in the form of Fig. 3a and the records of sediment production in the European Alps and the Himalaya (Fig. 3b) is striking, and supports the notion that a late deglacial pulse of sediment generation also occurred in these mountainous regions.

Figure 4 investigates the impact on the marine Sr isotope budget of the suggestion that present-day Sr fluxes are out of steady state (see Supplementary Information for model details). In addition to the changes in the silicate weathering flux detailed above, our calculations also take into account well-documented effects arising from the incongruent release of radiogenic Sr early in the weathering process (see Supplementary Information for details). As with Pb isotopes<sup>7,8</sup>, young soils preferentially release highly radiogenic Sr during initial weathering<sup>6</sup>, and the co-release of the radiogenic Pb (ref. 8) and Sr is supported by granitoid leaching experiments (Supplementary Information).

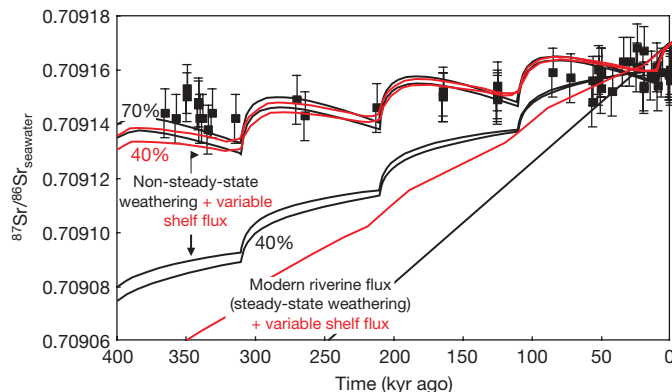
Figure 4 compares outputs of models of deglacial weathering with high resolution measurements of Sr isotopes in foraminiferal carbonate over the past 400 kyr (ref. 25). The isotope record places tight limits on the Sr isotopic evolution of sea water and the amplitude of



**Figure 3 | Chemical and physical weathering rates over the last glacial cycle.** **a**, Model of instantaneous silicate chemical weathering rate (black curve) in the 25–30% (ref. 19) of the extra-Antarctic continent that was glaciated at some point over the last glacial cycle, using the relationships between soil age and weathering rate from Fig. 1 and ice-sheet coverage forced by marine oxygen isotope data (see Supplementary Information for details). At deglaciations, for example, the last starting approximately 18 kyr ago, weathering rates rise rapidly as pristine soil parent material is exposed from underneath the retreating ice-sheet. The rate then decays away during the subsequent interglacial as this material ages (Fig. 2), but present-day weathering rates are still modelled to be about a factor of two higher than the rate over an entire glacial cycle (lower, darker-grey band). The ranges in both this long-term and the instantaneous weathering rates (the thickness of the bands) are given by assuming the range of exponential dependencies of weathering rate on time seen in Fig. 2. The upper, lighter-grey band gives the average (weighted according to drainage basin area) of the silicate weathering rates estimated for the Mackenzie and St Lawrence river basins at the present day<sup>23</sup>. **b**, Production of new sediment in peri-Alpine basins over the last glacial cycle (solid curve, right axis, expressed as an equivalent denudation rate for the mountain belt as in the original paper<sup>13</sup>) and sediment storage in the Ganges-Brahmaputra (G-B) floodplain (dashed line, left axis<sup>12</sup>) over the last deglaciation. These changes in sediment supply would produce pulses in chemical weathering rate very similar to those modelled for the high-latitude glaciated regions in **a**.

variations. Figure 4 demonstrates once again that if the modern riverine flux of Sr and its isotopes is representative of the long-term flux, then sea water evolves in  $^{87}\text{Sr}/^{86}\text{Sr}$  much more rapidly than the data allow. The modelled rate of evolution in seawater Sr isotopes matches data more closely, given maximum allowable changes in the flux of unradiogenic Sr from shelf carbonates during the sea-level low-stands of glacial periods (see Supplementary Information for details), but is still a factor of six too fast. However, if about 70% of the silicate weathering flux is affected by non-steady-state processes, the modelled evolution of sea water matches the data over the long term. Alternatively, a lower limit of 40% of the silicate chemical weathering flux affected by the processes is permitted if maximum increases in the unradiogenic flux from carbonates during glacial sea-level low-stands are allowed. Our weathering scenario produces cyclicity in the Sr isotopic composition of sea water, such that deglaciations are characterized by rapid ( $\sim 10$  kyr) increases in  $^{87}\text{Sr}/^{86}\text{Sr}$ , followed by slow ( $\sim 90$  kyr) reductions. The cyclicity generated, however, is not inconsistent with the best data we currently have for Sr isotopes in the oceans over the past 400 kyr (ref. 25).

The above ideas, successful for Sr, have implications for the oceanic budgets of other elements, whose quantitative validity might be tested in the future. First, our weathering scenario could explain similar imbalances in the marine geochemical budgets of other elements with



**Figure 4 | Sr behaviour during weathering and the oceanic mass balance.** The evolution of  $^{87}\text{Sr}/^{86}\text{Sr}$  in the oceans from records in marine carbonate<sup>25</sup> (filled squares; error bars represent 95% confidence intervals) compared to those produced by model simulations. The two lines labelled 'modern riverine flux (steady-state weathering)' assume a Sr flux and isotopic composition from the continental crust as given by the modern riverine flux with (red) or without (black) a maximum increased flux of unradiogenic Sr from carbonate shelves during glacial sea-level low-stands<sup>30</sup>, and suggest an evolution that is 6–8 times faster than observed. On the other hand, if 40–70% of the silicate weathering flux is subject to non-steady-state weathering caused by variable substrate ages, the models are consistent with data. In the models (see Supplementary Information for details), the specified portion of the silicate weathering flux is affected by variable substrate ages through time (Fig. 3), such that weathering rates are reset to high values by young substrates produced at deglaciations. The weathering flux of Sr is also affected by substrate-age-dependent incongruent weathering of Sr isotopes. The two curves for each percentage are for different degrees of incongruent weathering of Sr isotopes at the present day, such that  $^{87}\text{Sr}/^{86}\text{Sr}$  weathered from modern soils is 1.002 and 1.004 times the bulk soil value (or 0.2% and 0.4% higher than the bulk soil; see ref. 6, main text and Supplementary Information).

long oceanic residence times. The modern riverine input of Mg, even when a secular increase in oceanic concentrations is taken into account<sup>4</sup>, is also much greater than the current best estimates of the outputs<sup>4,17</sup>. As a consequence of its long ocean residence time (10 Myr) and the dominance of slow pre-Quaternary rates in the long-term average<sup>19</sup>, the modern supply of Mg to the oceans is even further out of step than Sr with the relevant long-term value. The assessment of the quantitative impact of this assertion is complicated by major uncertainty in the fraction of the riverine Mg budget that derives from silicate weathering (21–54%; ref. 18) and additional uncertainties over the long-term magnitude of key output fluxes, such as those to dolomite or to basalts<sup>4,17</sup>. Such uncertainties at present preclude the use of the Mg budget to test the hypothesis of non-steady-state weathering, but our models do represent a target for improved information on the global budget of Mg that is likely to emerge in the near future from its nascent isotope system<sup>17</sup>.

Second, for elements with an isotope system and whose residence times are sufficiently short to resolve glacial–interglacial changes in the weathering flux, we predict discernible glacial–interglacial cyclicity in their oceanic isotope geochemistry.

In addition to Pb, the oceanic behaviour of which has already been shown to be consistent with our ideas<sup>8</sup>, Os is another element for which glacial–interglacial cycles in its seawater isotopic composition are well established<sup>15</sup>. These changes are attributed to changes in the flux of radiogenic Os from the continents resulting from low weathering rates during glacial periods. However, modelling of Os isotopes purely in terms of lower weathering fluxes during glacial periods has shown that the glacial–interglacial changes can only be explained with an oceanic residence time for Os near 3 kyr (ref. 15), inconsistent with the much longer residence time (35–50 kyr) suggested by the small contrast in the Os isotopic compositions of different ocean basins. The only way to explain the glacial–interglacial changes with the longer residence

times is to invoke a transient Os spike at deglaciations due to rapid weathering of glacial till<sup>15</sup>, in direct congruence with the processes invoked here. Furthermore, previous work on the marine Os cycle has not considered possible changes in the isotopic composition of the riverine flux. Os is affected by exactly the same processes that affect the isotopic composition of Sr released in the weathering environment, with young soils releasing highly radiogenic Os and freshwater Fe-Mn nodules in post-glacial lakes containing extreme Os isotope compositions<sup>26</sup>.

Though not directly addressed here, our ideas invoking faster deglacial weathering rates would drive atmospheric CO<sub>2</sub> downwards by 10–20 p.p.m. when ice cores record a rise at deglaciations<sup>27</sup>. A full assessment of the magnitude of this phenomenon must await a comprehensive carbon cycle model, including a determination of the timing of predicted pulses in silicate weathering in the glaciated regions in response to ice-sheet retreat early in a deglaciation versus a later pulse in response to a strong early Holocene monsoon. A corollary to the ideas presented here (see also ref. 19) is that long-term CO<sub>2</sub> drawdown in an ice-house world such as the Quaternary—with cycles of physical (glacial) grinding making material available for subsequent intense chemical (interglacial) weathering—is faster than in the absence of these climate changes. The implication is that the CO<sub>2</sub>–weathering feedback is weakened (and perhaps even changed in sign) during periods like the Quaternary, perhaps contributing to the maintenance of ice-house conditions once they are initiated.

Received 17 July 2008; accepted 22 January 2009.

- Elderfield, H. & Schultz, A. Mid-ocean hydrothermal fluxes and the chemical composition of the ocean. *Annu. Rev. Earth Planet. Sci.* **24**, 191–224 (1996).
- Galy, A., France-Lanord, C. & Derry, L. A. The strontium isotopic budget of Himalayan rivers in Nepal and Bangladesh. *Geochim. Cosmochim. Acta* **63**, 1905–1925 (1999).
- Davis, A. C., Bickle, M. J. & Teagle, D. A. H. Imbalance in the oceanic strontium budget. *Earth Planet. Sci. Lett.* **211**, 173–187 (2003).
- Holland, H. D. Sea level, sediments and the composition of seawater. *Am. J. Sci.* **305**, 220–239 (2005).
- Taylor, A. & Blum, J. D. Relation between soil age and silicate weathering rates determined from the chemical evolution of a glacial chronosequence. *Geology* **23**, 979–982 (1995).
- Blum, J. D. & Erel, Y. Rb–Sr isotope systematics of a granitic soil chronosequence: The importance of biotite weathering. *Geochim. Cosmochim. Acta* **61**, 3193–3204 (1997).
- Harlavan, Y., Erel, Y. & Blum, Y. D. Systematic changes in lead isotopic composition with soil age in glacial granitic terrains. *Geochim. Cosmochim. Acta* **62**, 33–46 (1998).
- Foster, G. L. & Vance, D. Negligible glacial-interglacial variation in continental chemical weathering rates. *Nature* **444**, 918–921 (2006).
- Dosseto, A., Bourdon, B. & Turner, S. P. Uranium-series isotopes in river materials: Insights into the timescales of erosion and sediment transport. *Earth Planet. Sci. Lett.* **265**, 1–17 (2008).
- Bell, M. & Laine, E. P. Erosion of the Laurentide region of North America by glacial and glaciofluvial processes. *Quat. Res.* **23**, 154–174 (1985).
- Thomas, M. F. & Thorp, M. B. Geomorphic response to rapid climatic and hydrologic change during the late Pleistocene and early Holocene in the humid and sub-humid tropics. *Quat. Sci. Rev.* **14**, 193–197 (1995).
- Goodbred, S. L. Jr & Kuehl, S. A. Enormous Ganges-Brahmaputra sediment discharge during strengthened early Holocene monsoon. *Geology* **28**, 1083–1086 (2000).
- Hinderer, M. Late Quaternary denudation of the Alps, valley and lake fillings and modern river loads. *Geodinam. Acta* **14**, 231–263 (2001).
- Hodell, D. A., Mueller, P. A., McKenzie, J. A. & Mead, G. A. Strontium isotope stratigraphy and geochemistry of the late Neogene ocean. *Earth Planet. Sci. Lett.* **92**, 165–178 (1989).
- Oxburgh, R., Pierson-Wickmann, A.-C., Reisberg, L. & Hemming, S. Climate-correlated variations in seawater <sup>187</sup>Os/<sup>188</sup>Os over the past 200,000 yr: Evidence from the Cariaco Basin, Venezuela. *Earth Planet. Sci. Lett.* **263**, 246–258 (2007).
- Palmer, M. R. & Edmond, J. M. The strontium isotope budget of the modern ocean. *Earth Planet. Sci. Lett.* **92**, 11–26 (1989).
- Tipper, E. T. et al. The magnesium isotope budget of the modern ocean: Constraints from riverine magnesium isotope ratios. *Earth Planet. Sci. Lett.* **250**, 241–253 (2006).
- Gaillardet, J., Dupré, B., Louvat, P. & Allègre, C. J. Global silicate weathering and CO<sub>2</sub> consumption rates deduced from the chemistry of large rivers. *Chem. Geol.* **159**, 3–30 (1999).
- Clark, P. U. et al. The middle Pleistocene transition: Characteristics, mechanisms and implications for long-term changes in atmospheric CO<sub>2</sub>. *Quat. Sci. Rev.* **25**, 3150–3184 (2006).
- Anderson, S. P. Biogeochemistry of glacial landscape systems. *Annu. Rev. Earth Planet. Sci.* **35**, 375–399 (2007).
- White, A. F. & Brantley, S. L. The effect of time on the weathering of silicate minerals: Why do weathering rates differ in the laboratory and the field? *Chem. Geol.* **202**, 479–506 (2003).
- Porder, S., Hilley, G. E. & Chadwick, O. A. Chemical weathering, mass loss, and dust inputs across a climate by time matrix in the Hawaiian Islands. *Earth Planet. Sci. Lett.* **258**, 414–427 (2007).
- Millot, R., Gaillardet, J., Dupré, B. & Allègre, C. J. The global control of silicate weathering rates and the coupling with physical erosion: New insights from rivers of the Canadian Shield. *Earth Planet. Sci. Lett.* **196**, 83–98 (2002).
- West, A. J., Galy, A. & Bickle, M. Tectonic and climatic controls on silicate weathering. *Earth Planet. Sci. Lett.* **235**, 211–228 (2005).
- Henderson, G. M., Martel, D. J., O’Nions, R. K. & Shackleton, N. J. Evolution of seawater <sup>87</sup>Sr/<sup>86</sup>Sr over the last 400 ka: The absence of glacial/interglacial cycles. *Earth Planet. Sci. Lett.* **128**, 643–651 (1994).
- Peucker-Ehrenbrink, B. & Blum, J. D. Re-Os isotope systematics and weathering of Precambrian crustal rocks: Implications for the marine osmium isotope record. *Geochim. Cosmochim. Acta* **62**, 3193–3203 (1998).
- Petit, J. R. et al. Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* **399**, 429–436 (1999).
- Teagle, D. A. H., Bickle, M. J. & Alt, J. C. Recharge flux to ocean-ridge black smoker systems: A geochemical estimate from ODP Hole 504B. *Earth Planet. Sci. Lett.* **210**, 81–89 (2003).
- Nielsen, S. G. et al. Hydrothermal fluid fluxes calculated from the isotopic mass balance of thallium in the ocean crust. *Earth Planet. Sci. Lett.* **251**, 120–133 (2006).
- Stoll, H. M. & Schrag, D. P. Effect of Quaternary sea-level cycles on Sr in seawater. *Geochim. Cosmochim. Acta* **62**, 1107–1118 (1998).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank H. Pälike, M. Palmer, C. Hawkesworth, T. Elliott, D. Schmidt, M. Gutjahr and M. Andersen for reading and improving an earlier version of this paper, and J. Phillips for help with a statistics problem. This work was done while D.V. held a Blaustein Visiting Research Professorship at Stanford.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to D.V. ([d.vance@bristol.ac.uk](mailto:d.vance@bristol.ac.uk)).



# Volcanic mesocyclones

Pinaki Chakraborty<sup>1</sup>, Gustavo Gioia<sup>2</sup> & Susan W. Kieffer<sup>1</sup>

**A strong volcanic plume consists of a vertical column of hot gases and dust topped with a horizontal ‘umbrella’<sup>1</sup>. The column rises, buoyed by entrained and heated ambient air, reaches the neutral-buoyancy level, then spreads radially to form the umbrella. In classical models of strong volcanic plumes, the plume is assumed to remain always axisymmetric and non-rotating. Here we show that the updraught of the rising column induces a hydrodynamic effect not addressed to date—a ‘volcanic mesocyclone’. This volcanic mesocyclone sets the entire plume rotating about its axis, as confirmed by an unprecedented analysis of satellite images from the 1991 eruption of Mount Pinatubo<sup>2–4</sup>. Destabilized by the rotation, the umbrella loses axial symmetry and becomes lobate in plan view, in accord with satellite records of recent eruptions on Mounts Pinatubo, Manam, Reventador, Okmok, Chaiten and Ruang. The volcanic mesocyclone spawns waterspouts<sup>5,6</sup> or dust devils<sup>6–8</sup>, as seen in numerous eruptions, and groups the electric charges about the plume to form the ‘lightning sheath’ that was so prominent in the recent eruption of Mount Chaiten. The concept of a volcanic mesocyclone provides a unified explanation for a disparate set of poorly understood phenomena in strong volcanic plumes<sup>5–10</sup>.**

On 12 June 1811, a volcanic vent surfaced out of the sea in the Azores archipelago. A volcanic column started to rise and was observed by a ship’s captain. According to the captain<sup>3</sup>, the column rotated on the water “like an horizontal wheel” and was accompanied by flashes of lightning which “continually issued from the densest part of the volcano.” Then the column “rolled off in large masses of fleecy clouds, gradually expanding themselves in a direction nearly horizontal, and drawing up to them a quantity of waterspouts.” Although the three features observed by the captain—rotation of the plume about its axis, lightning close to the column, and waterspouts (or dust devils) eccentric to the column—have variously been noted over the years<sup>6–10</sup>, the captain’s report appears to be the only account of a volcanic plume in which all of these features have been noted together. What has not been noted is that these features are characteristic of a meteorological phenomenon seemingly unrelated to volcanic plumes: the cyclonically rotating columnar vortex—or mesocyclone—of a tornadic thunderstorm<sup>11</sup>. Here we introduce the concept of a volcanic mesocyclone. We start by providing direct evidence of the rotation attendant on a volcanic mesocyclone.

Consider Mount Pinatubo’s eruption of 15 June 1991, the only large volcanic eruption for which there exists a satellite record suitable for the analysis that follows. In that eruption, the umbrella was documented by the Geostationary Meteorological Satellite (GMS), in the form of hourly images in plan view (Supplementary Fig. 1). From these images, Holasek *et al.*<sup>2</sup> have extracted the contours of the edge of the umbrella at four stages, 60 min apart from one another (Fig. 1a). We have processed these contours to trace the horizontal displacement of the centre of the umbrella (if any), the rotation of the edge of the umbrella about its centre (if any), and the radial expansion of the edge of the umbrella (Fig. 1b). Our calculations show that the edge of Pinatubo’s umbrella rotated about its centre<sup>2,12</sup>. The rate of

rotation decayed from  $\sim 0.5 \text{ rad h}^{-1}$  to  $\sim 0.2 \text{ rad h}^{-1}$  over a period of 2 h (Fig. 1c). Further, a comparison of the second and third contours of Fig. 1a indicates that at some time between 14:41 and 15:41 Philippine Daylight Time (PDT), the edge of the umbrella became wavy—the umbrella took a lobate shape in plan view. We estimate the number of lobes,  $N = 5$ , from the contour for 16:41 PDT, by which time the amplitude of the lobes attained a value comparable to the wavelength. The number of lobes may be verified in the satellite image of Fig. 1d.

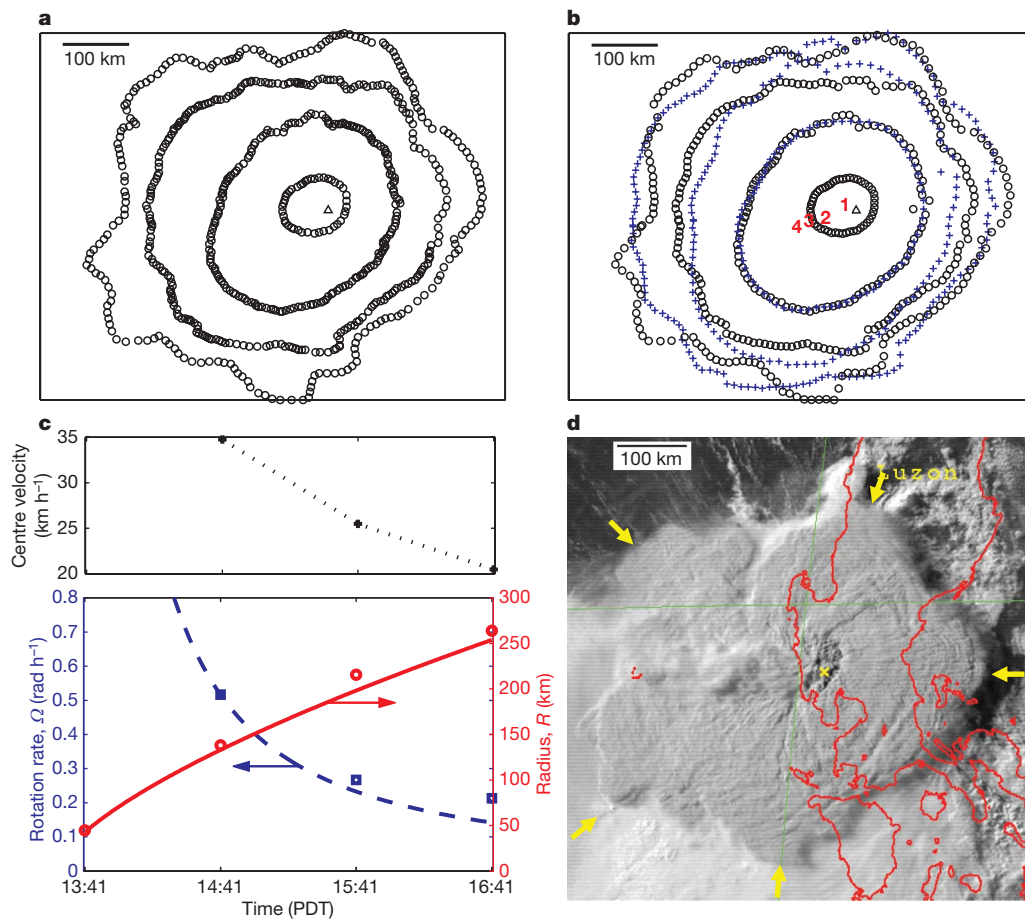
Why was Pinatubo’s umbrella rotating, and why did it become lobate? We argue that (1) the rotation of an umbrella is inherited from the column, which in turn rotates because of the spontaneous development of a volcanic mesocyclone, and (2) the lobateness of an umbrella is a direct consequence of the rotation, which makes the umbrella unstable and contorts its edge.

To understand why a volcanic plume rotates, consider the development of the mesocyclone of a tornadic thunderstorm. When a vigorous updraught rises in the presence of the horizontal vortex tubes associated with the strong shear layer of an ambient wind, the updraught entrains, tilts and stretches these vortex tubes to produce a pair of counter-rotating vertical vortices<sup>11,13</sup>. Of these counter-rotating vertical vortices, the cyclonically rotating vortex (that is, the vortex that rotates anticlockwise in the Northern Hemisphere or clockwise in the Southern Hemisphere) is preferentially enhanced by the turning of the wind-shear vector with height—a subtle effect that is ultimately caused by the rotation of the Earth<sup>11,13</sup>. We propose that similar processes occur in the updraught of a volcanic column, resulting in a volcanic mesocyclone that ultimately sets the entire plume rotating (Fig. 2a). (For further discussion, see Supplementary Information.)

Besides a strong shear layer (of ambient winds), there are additional sources of localized vorticity for a volcanic mesocyclone. The periphery of a strong volcanic column is lined with the eddies of a Kelvin–Helmholtz instability, which is driven by the shear between the column and the surrounding atmosphere<sup>1</sup>. These eddies form horizontal vortex rings (Fig. 2b) that are frequently present in strong volcanic plumes (where the updraught velocities are as high as  $600 \text{ m s}^{-1}$  in the gas-thrust region, and  $200 \text{ m s}^{-1}$  in the convective region<sup>1</sup>), but not in thunderstorms (where the updraught velocities are of the order of  $10 \text{ m s}^{-1}$ , even in supercells<sup>14</sup>). During entrainment of ambient air, these vortex rings (as well as the intense turbulent eddies that fill a volcanic column) may act as additional sources of vorticity in the formation of a volcanic mesocyclone (Fig. 2b). Thus a volcanic mesocyclone is likely to be a more robust phenomenon than its counterpart in a thunderstorm.

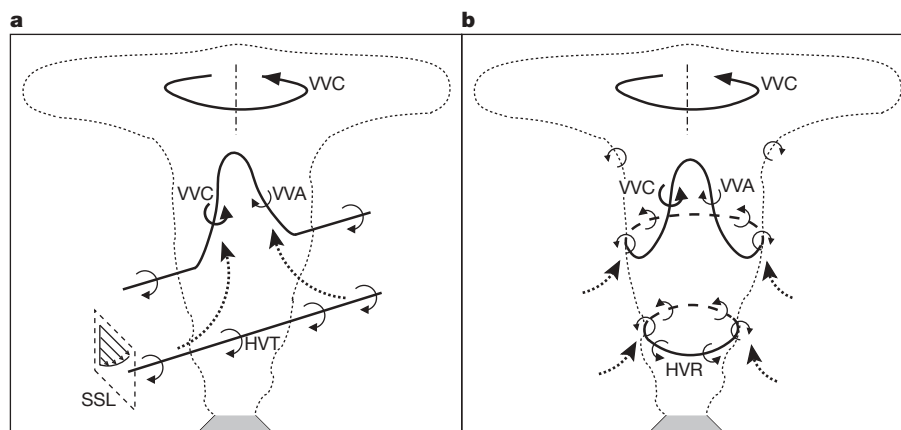
A number of mechanisms<sup>15,16</sup> might be adduced to explain why a rotating umbrella loses axial symmetry and becomes lobate in plan view (see Supplementary Information). Here we discuss one such mechanism. We assume that the edge of the umbrella is initially circular in plan view and rotates with angular velocity  $\Omega$ . Further, the thickness of the umbrella is negligible compared to the radius  $R$ , so that flow in the umbrella is quasi-two-dimensional. Under these

<sup>1</sup>Department of Geology, <sup>2</sup>Department of Mechanical Science & Engineering, University of Illinois, Urbana, Illinois 61801, USA.



**Figure 1 | Analysis of Pinatubo's umbrella (15 June 1991).** **a**, Contours of the edge of Pinatubo's umbrella at hourly intervals from 13:41 PDT (the first, smallest contour) to 16:41 PDT (the fourth, largest contour); Mount Pinatubo is indicated by a triangle. (PDT, Philippine Daylight Time.) After ref. 2. **b**, Black contours (open circles) are those of **a** for 13:41, 14:41, 15:41 and 16:41 PDT. Red numbers 1 to 4 indicate the calculated centres of rotation (see below) of the black contours, starting with the black contour for 13:41, the centroid of which we take to be the centre of rotation. Blue contours (crosses) are processed contours for 14:41, 15:41 and 16:41 PDT; to obtain the processed contour for 15:41, say, we displace, rotate, and expand the black contour for 14:41 about its centre of rotation for best fit with the

black contour for 15:41. (See Supplementary Information for details.) **c**, Top panel, the velocity of the centre of rotation versus time, from the analysis of satellite images. Bottom panel, the radius of the umbrella,  $R$ , versus time (red open circles) and the rate of rotation,  $\Omega$ , versus time (open blue squares), from the analysis of satellite images. Solid lines are theoretical predictions for  $R(t)$  (from ref. 2) and for  $\Omega(t)$  (from imposing conservation of angular momentum:  $\Omega(t)R(t)^2 = \text{constant} = \Omega(t=14:41)R(t=14:41)^2$ ). **d**, Satellite image of the lobate umbrella at 16:41 PDT<sup>2-4</sup>. Yellow arrows, location of lobes; yellow cross, location of Mount Pinatubo; red outlines, the Philippine islands including Luzon.



**Figure 2 | Diagrams to illustrate the formation of a volcanic mesocyclone starting from different sources of vorticity (not to scale).** We consider two sources: **a**, the horizontal vortex tubes (HVT) associated with the strong shear layer (SSL) of an ambient wind, and **b**, the horizontal vortex rings (HVR) associated with the Kelvin-Helmholtz instability that develops between the rising column and the surrounding atmosphere. The horizontal

vortex tubes of **a** and the horizontal vortex rings of **b** are entrained, tilted and stretched by the updraught to produce counter-rotating vertical vortices in the updraught of the volcanic plume. Of the counter-rotating vertical vortices, the cyclonic vortex (VVC) becomes dominant relative to the anticyclonic vortex (VVA).

assumptions and in a frame of reference that rotates with the edge of the umbrella, the fluid in the umbrella is subjected to two radial body forces near the edge of the umbrella: a centrifugal force  $\Omega^2 R$  and a Coriolis force  $\pm 2\Omega u$ , where  $u$  is the magnitude of the turbulent (fluctuating) fluid velocity. Near the edge of the umbrella  $u$  is tangent to the edge, and its magnitude is set by the velocity of the largest turbulent eddies, which is about  $10^{-2}$  times the characteristic velocity of the flow<sup>17–19</sup>. As the characteristic velocity of the flow is  $\Omega R$ ,  $u \approx 10^{-2} \Omega R$ , and the Coriolis force has a magnitude of  $\sim 2 \times 10^{-2} \Omega^2 R$ . We conclude that the dominant body force near the edge of the umbrella is the centrifugal force. It is directed radially outward and has a magnitude of  $\Omega^2 R$ .

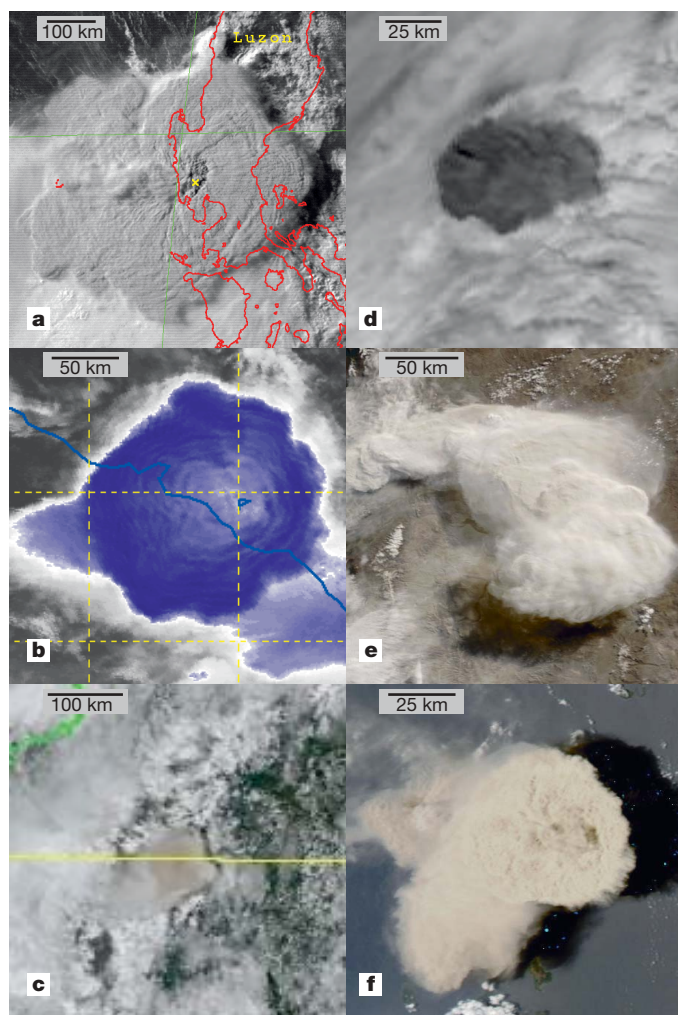
As the umbrella expands, the fluid near its edge cools and thus becomes denser than the surrounding atmosphere. This density contrast coupled with the centrifugal force (which is directed radially outward from the dense edge of the umbrella towards the less dense atmosphere) triggers a baroclinic instability—a turbulent, centrifugal form of the Rayleigh–Taylor instability. (In a turbulent Rayleigh–Taylor instability<sup>20</sup>, diffusion is governed by eddy viscosity<sup>1,21,22</sup> as opposed to molecular viscosity.) By following steps similar to those outlined in ref. 20, we derive a mathematical relation between  $\Omega$ ,  $N$ , and the characteristic time of the instability,  $\tau$ , where  $\tau$  represents the

time required for the amplitude of the lobes to attain a value comparable to the wavelength (see Supplementary Information for details):

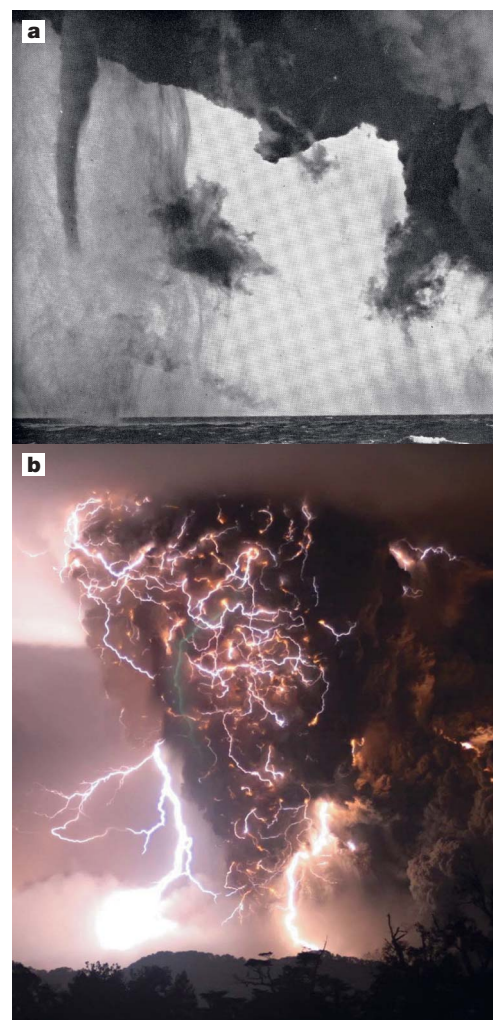
$$\tau \Omega \sqrt{N} = 1 \quad (1)$$

From our analysis of satellite images of Pinatubo (Fig. 1a), we concluded that at some time between 14:41 and 15:41 PDT, the Pinatubo umbrella became unstable and developed  $N = 5$  lobes. Therefore, the angular velocity was  $\Omega \approx 0.4 \text{ rad h}^{-1}$  when the instability was triggered (at the midpoint between 14:41 and 15:41 PDT (Fig. 1c)). From our analysis of satellite images of Pinatubo, we also concluded that at 16:41 PDT, about 1.5 h after the instability was triggered, the amplitude of the lobes attained a value comparable to the wavelength. It follows that we would expect equation (1) to give  $\tau \approx 1.5 \text{ h}$ ; substituting  $\Omega = 0.4 \text{ rad h}^{-1}$  and  $N = 5$  in equation (1), we get  $\tau = 1.1 \text{ h}$ , in reasonable accord with expectation.

As the conditions for the development of a volcanic mesocyclone are commonly realized in strong volcanic plumes, we predict that the umbrellas of such plumes usually rotate and become lobate. This prediction may be confirmed in satellite records (Fig. 3). We propose that the rotation of an umbrella and the attendant lobateness are distinctive signatures of a volcanic mesocyclone. Other signatures are the spawning of tornadoes (waterspouts on water and dust devils on land) and the formation of sheaths of lightning. We address the spawning of tornadoes first.



**Figure 3 | Satellite images of lobate umbrellas.** **a**, Pinatubo (Philippines) on 15 June 1991, 16:41 local time (diameter  $\sim 540 \text{ km}$ ); **b**, Manam (Papua New Guinea) on 27 January 2005, 15:35 UTC (180 km); **c**, Reventador (Ecuador) on 3 November 2002, 15:10 UTC (85 km); **d**, Okmok (Alaska, USA) on 12 July 2008, 20:43 UTC ( $\sim 65 \text{ km}$ ); **e**, Chaiten (Chile) on 6 May 2008, 15:05 UTC ( $\sim 115 \text{ km}$ ); and **f**, Ruang (Indonesia) on 25 September 2002, 04:50 UTC ( $\sim 60 \text{ km}$ ). Data sources: GMS (**a**); NASA MODIS (**b**, **c**, **e**, **f**); NOAA/AVHRR (**d**).



**Figure 4 | Secondary signatures of a volcanic mesocyclone.** **a**, Waterspouts spawned during the eruption of Surtsey volcano on 14 November 1963 (ref. 6). **b**, A lightning sheath covers the volcanic column from Mount Chaiten on 3 May 2008. (Used with permission from Landov Media.)



In a tornadic thunderstorm, the key elements that interact (in ways imperfectly understood) to yield tornadoes eccentric to the axis of the thunderstorm are “a buoyant updraft, rainy downdrafts, and a deep, mesocyclonic vortex (preexisting vertical vorticity)”<sup>13</sup>. We propose that in a volcanic eruption the same elements—updraught (which has been extensively studied<sup>1</sup>), downdraughts (which have received little attention<sup>23,24</sup>) and the volcanic mesocyclone (which is the subject of this work)—yield the tornadic structures, such as dust devils and waterspouts that are frequently observed in strong volcanic plumes<sup>5–8</sup> (for example, Fig. 4a).

In connection with lightning<sup>7,9,10</sup> in strong volcanic plumes, we discuss supercells, a type of thunderstorm that might be the closest analogue to strong volcanic plumes. A supercell is a thunderstorm with an intense mesocyclone and a highly organized internal structure<sup>14</sup>. Within the swift updraught of a supercell there is insufficient time for a sizable amount of precipitation to form, grow and gain charge via collisions<sup>25,26</sup>. As a result, lightning remains minimal within the updraught of a supercell, and the core of the updraught has been termed a ‘lightning hole’<sup>25–27</sup>. It has recently been proposed that in supercells the mesocyclone pulls the precipitation radially outwards from the core of the updraught, gathering the precipitation over the periphery of the updraught, where it sets up the formation of a ‘lightning sheath’ (or ‘ring’<sup>28</sup>) around the updraught<sup>29</sup>. Remarkably, photographs from the recent eruption of Chaiten (for example, Fig. 4b) show the surface of the volcanic column prominently coated in a layer of lightning, which we identify as a lightning sheath.

We have argued that strong volcanic plumes are accompanied by volcanic mesocyclones. Signatures of a volcanic mesocyclone include a rotating plume, a lobate umbrella, the spawning of tornadoes, and the formation of a lightning sheath and hole. These signatures, which remain unaccounted for in current models, must be present in most strong volcanic plumes, even though they may be obscured by distorting winds and other phenomena unrelated to the volcanic eruption. We hope that these signatures will be the subject of future remote sensing observations and field work on volcanic plumes. Satellite images at intervals of a few minutes would make it possible to trace the evolution of umbrellas in detail. The structure and dynamics of volcanic mesocyclones, as well as the presence of lightning sheaths and holes, might be verified using respectively Doppler radar<sup>30</sup> and lightning mapping arrays<sup>10</sup>, two technologies that have been scarcely used in volcanology. Last, we hope that the concept of a volcanic mesocyclone will help us forecast, and alleviate, the impact of volcanic eruptions.

Received 16 October 2008; accepted 11 February 2009.

1. Sparks, R. S. J. *et al.* *Volcanic Plumes* (Wiley and Sons, 1997).
2. Holasek, R. E., Self, S. & Woods, A. W. Satellite observations and interpretations of the 1991 Mount Pinatubo eruption plumes. *J. Geophys. Res.* **101**, 27635–27655 (1996).
3. Oswalt, J. S., Nichols, W. & O'Hara, J. F. in *Fire and Mud: Eruptions and Lahars of Mount Pinatubo, Philippines* (eds Newhall, C. G. & Punongbayan, R. S.) 625–636 (Philippine Institute of Volcanology and Seismology and University of Washington Press, 1996).
4. Self, S., Zhao, J., Holasek, R. E., Torres, R. C. & King, A. in *Fire and Mud: Eruptions and Lahars of Mount Pinatubo, Philippines* (eds Newhall, C. G. & Punongbayan, R. S.) 1089–1115 (Philippine Institute of Volcanology and Seismology and University of Washington Press, 1996).

5. Tillard, S. A narrative of the eruption of a volcano in the sea off the island of St. Michael. *Phil. Trans. R. Soc. Lond. B* **102**, 152–158 (1812).
6. Thorarinsson, S. & Vonnegut, B. Whirlwinds produced by the eruption of Surtsey volcano. *Bull. Am. Meteorol. Soc.* **45**, 440–444 (1964).
7. Anderson, R. *et al.* Electricity in volcanic clouds: investigations show that lightning can result from charge-separation processes in a volcanic crater. *Science* **148**, 1179–1189 (1965).
8. Stothers, R. B. The great Tambora eruption in 1815 and its aftermath. *Science* **224**, 1191–1198 (1984).
9. Mather, T. A. & Harrison, R. G. Electrification of volcanic plumes. *Surv. Geophys.* **27**, 387–432 (2006).
10. Thomas, R. J. *et al.* Electrical activity during the 2006 Mount St. Augustine volcanic eruptions. *Science* **315**, 1097 (2007).
11. Klemp, J. B. Dynamics of tornadic thunderstorms. *Annu. Rev. Fluid Mech.* **19**, 369–402 (1987).
12. Baines, P. G. & Sparks, R. S. J. Dynamics of giant volcanic ash clouds from supervolcanic eruptions. *Geophys. Res. Lett.* **32**, L24808, doi:10.1029/2005GL024597 (2005).
13. Davis-Jones, R., Trapp, R. J. & Bluestein, H. B. in *Severe Convective Storms* Ch. 5 (ed. Doswell, C. A. III) 167–221 (Meteorological Monographs Vol. 28, American Meteorological Society, 2001).
14. Emanuel, K. *Atmospheric Convection* (Oxford Univ. Press, 1994).
15. Griffiths, R. W. Gravity currents in rotating systems. *Annu. Rev. Fluid Mech.* **18**, 59–89 (1986).
16. Linden, P. F. in *Rotating Fluids in Geophysical and Industrial Applications* (ed. Hopfinger, E. J.) 99–123 (Springer, 1992).
17. Tennekes, H. & Lumley, J. L. *A First Course in Turbulence* (MIT Press, 1972).
18. Landau, L. D. & Lifshitz, E. M. *Course of Theoretical Physics* Vol. 6, *Fluid Mechanics* (Elsevier Academic Press, 2000).
19. Gioia, G. & Chakraborty, P. Turbulent friction in rough pipes and the energy spectrum of the phenomenological theory. *Phys. Rev. Lett.* **96**, 044502 (2006).
20. Chakraborty, P., Gioia, G. & Kieffer, S. Volcan Reventador's unusual umbrella. *Geophys. Res. Lett.* **33**, L05313, doi:10.1029/2005GL024915 (2006).
21. Valentine, G. A. & Wohletz, K. H. Numerical models of plinian eruption columns and pyroclastic flows. *J. Geophys. Res.* **94**, 1867–1887 (1989).
22. Pedlosky, J. *Geophysical Fluid Dynamics* (Springer, 1979).
23. Hoblitt, R. P. Was the 18 May 1980 lateral blast at Mt. St. Helens the product of two explosions? *Phil. Trans. R. Soc. Lond. A* **358**, 1639–1661 (2000).
24. Tupper, A., Oswalt, J. S. & Rosenfeld, D. Satellite and radar analysis of the volcanic-cumulonimbi at Mt Pinatubo, Philippines, 1991. *J. Geophys. Res.* **110**, D09204, doi:10.1029/2004JD005499 (2005).
25. MacGorman, D. R. *et al.* The electrical structure of two supercell storms during STEPS. *Mon. Weath. Rev.* **133**, 2583–2607 (2005).
26. Wiens, K. C., Rutledge, S. A. & Tessendorf, S. A. The 29 June 2000 supercell observed during STEPS, part II: Lightning and charge structure. *J. Atmos. Sci.* **62**, 4151–4177 (2005).
27. Krehbiel, P. R. *et al.* GPS-based mapping system reveals lightning inside storms. *Eos* **81**, 21–25 (2000).
28. Payne, C. *The Evolution of a Lightning Hole during the 29–30 May 2004 HP Supercell during TELEX*. M.Sc. thesis, Univ. Oklahoma (2008).
29. Bruning, E. *Charging Regions, Regions of Charge, and Storm Structure in a Partially Inverted Polarity Supercell Thunderstorm*. Ph.D. thesis, Univ. Oklahoma (2008).
30. Dubosclard, G. *et al.* First testing of a volcano Doppler radar (VOLDORAD) at Mount Etna, Italy. *Geophys. Res. Lett.* **26**, 3389–3392 (1999).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We acknowledge discussions with A. Tupper, D. MacGorman, S. Nesbitt and M. Fromm. We thank the members of the ‘volcanic clouds’ discussion list for help with satellite images. P.C. and S.W.K. acknowledge support through the latter’s Walgreen Chair funds and NSF\EAR grant 06-09712 (S. Esperanza, programme director). G.G. acknowledges support through NSF\DMR grant 06-04435 (W. Fuller-Mora, programme director).

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to P.C. ([chakrabo@illinois.edu](mailto:chakrabo@illinois.edu)).

# Specific synapses develop preferentially among sister excitatory neurons in the neocortex

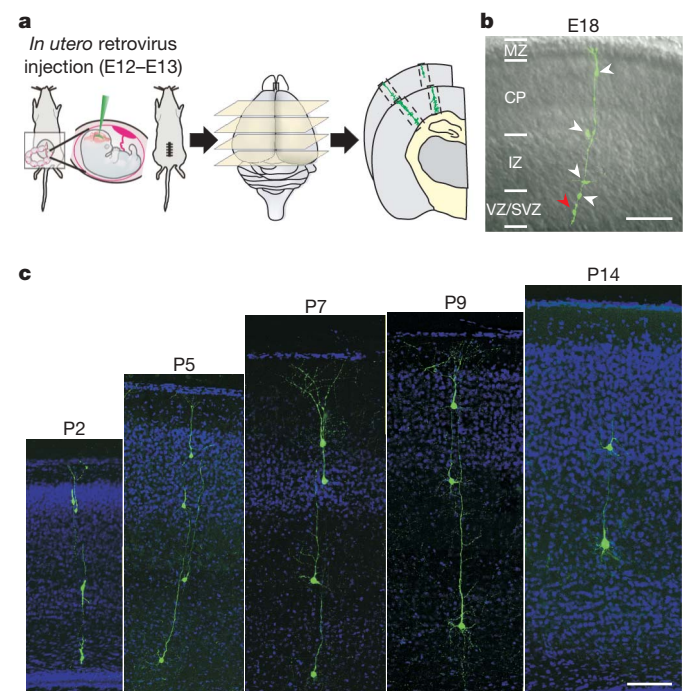
Yong-Chun Yu<sup>1</sup>, Ronald S. Bultje<sup>1,2</sup>, Xiaoqun Wang<sup>1</sup> & Song-Hai Shi<sup>1</sup>

Neurons in the mammalian neocortex are organized into functional columns<sup>1,2</sup>. Within a column, highly specific synaptic connections are formed to ensure that similar physiological properties are shared by neuron ensembles spanning from the pia to the white matter. Recent studies indicate that synaptic connectivity in the neocortex is sparse and highly specific<sup>3–8</sup> to allow even adjacent neurons to convey information independently<sup>9–12</sup>. How this fine-scale microcircuit is constructed to create a functional columnar architecture at the level of individual neurons largely remains a mystery. Here we investigate whether radial clones of excitatory neurons arising from the same mother cell in the developing neocortex serve as a substrate for the formation of this highly specific microcircuit. We labelled ontogenetic radial clones of excitatory neurons in the mouse neocortex by *in utero* intraventricular injection of enhanced green fluorescent protein (EGFP)-expressing retroviruses around the onset of the peak phase of neocortical neurogenesis. Multiple-electrode whole-cell recordings were performed to probe synapse formation among these EGFP-labelled sister excitatory neurons in radial clones and the adjacent non-siblings during postnatal stages. We found that radially aligned sister excitatory neurons have a propensity for developing unidirectional chemical synapses with each other rather than with neighbouring non-siblings. Moreover, these synaptic connections display the same interlaminar directional preference as those observed in the mature neocortex. These results indicate that specific microcircuits develop preferentially within ontogenetic radial clones of excitatory neurons in the developing neocortex and contribute to the emergence of functional columnar micro-architectures in the mature neocortex.

Recent studies have demonstrated that radial glial cells are the major neuronal progenitors in the developing neocortex<sup>13–17</sup>. In addition to their well-characterized role in guiding the radial migration of post-mitotic neurons<sup>18</sup>, radial glial cells divide asymmetrically to self-renew and give rise to neurons. Consecutive asymmetric cell divisions of individual radial glial cells produce several clonally related neurons that migrate radially into the cortical plate. This results in a columnar arrangement of neocortical neurons—the ontogenetic radial clone<sup>13,14,19,20</sup>. It has previously been suggested that ontogenetic columns become the basic processing unit in the adult cortex<sup>19</sup>; however, this assertion is largely on the basis of their anatomical similarity, that is, the vertical arrangement of neurons. In fact, virtually nothing is known about the functional development of ontogenetic radial clones in the neocortex.

To study the functional development of ontogenetic radial clones in the developing neocortex, we injected EGFP-expressing retroviruses through the uterus into the ventricle of mouse embryos around 12 to 13 days after conception (embryonic day 12 to 13, E12–E13; Fig. 1a)—the period during which the peak phase of neocortical neurogenesis

begins. After injection, the uterus was placed back to allow the embryos to come to term. Previous studies have shown that about half of the retroviral integration events occur in the self-renewing daughter cells<sup>21</sup>, resulting in the labelling of neocortical progenitor cells and their radially aligned clonal progeny<sup>13,20</sup>. Similarly, we found that intraventricular injection of a low titre EGFP-expressing retrovirus at E12–E13 reliably labelled the radial clones of cells (Fig. 1b, c), while also labelling some scattered cells (data not shown)<sup>22,23</sup>, throughout the developing neocortex. In this study we focused our analysis on the isolated radial clones that consisted of sister cells (Fig. 1 and Supplementary Fig. 1).



**Figure 1 | Morphological development of ontogenetic radial clones of excitatory neurons in the neocortex.** **a**, Labelling of ontogenetic radial clones in the developing neocortex with intraventricular injection of EGFP-expressing retroviruses. **b**, An EGFP-expressing ontogenetic radial clone observed at E18 with characteristic features: one cell resembles a bipolar radial glial cell (red arrowhead) and four other cells distribute radially along the radial glial fibre (white arrowheads). CP, cortical plate; IZ, intermediate zone; MZ, marginal zone; SVZ, subventricular zone; VZ, ventricular zone. Scale bar, 100  $\mu$ m. **c**, EGFP-expressing radial clones observed in individual brain sections at postnatal stages. Brain sections were counterstained with the DNA dye 4,6-diamidino-2-phenylindole (DAPI, blue). Scale bar, 100  $\mu$ m.

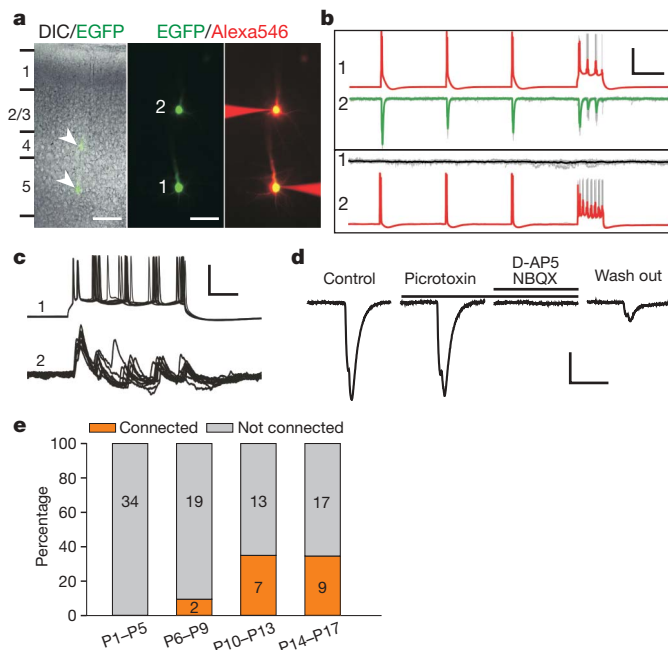
<sup>1</sup>Developmental Biology Program, Memorial Sloan Kettering Cancer Centre, 1275 York Avenue, <sup>2</sup>Department of Pharmacology, Weill Medical College of Cornell University, 445 East 69th Street, New York, New York 10065, USA.

In brain sections at various embryonic stages we frequently observed individual radial clones that comprised a radial glial mother cell (Fig. 1b, red arrowhead) and a cluster of further cells (~4–6 cells per clone at E18; Fig. 1b, white arrowheads), all of which were in contact with the radial glial fibre of the mother radial glial cell. As development progressed into postnatal stages, the frequency of radial clones observed in individual brain sections decreased. Nonetheless, we often found isolated radial clones with two or more cells in individual sections that did not have any scattered cells in close proximity (Fig. 1c and Supplementary Fig. 1), suggesting that the cells in these radial clones are lineage-related sisters. Nearly all of the cells in the labelled radial clones had the morphological features of an excitatory neuron—a radially oriented pyramid-shaped cell body and a major apical process that progressively branched (Fig. 1c) and harboured dendritic spines (data not shown) as they developed.

We proceeded to perform whole-cell patch clamp recordings to assess the physiological development of these ontogenetic radial clones in the developing neocortex during postnatal stages (Supplementary Fig. 2). As time progressed, the input resistance of neurons in the radial clone decreased (data not shown) and their resting membrane potential became progressively more hyperpolarized (Supplementary Fig. 2a, b), indicating that there is a maturation of membrane conductance. Furthermore, as development progressed the threshold for firing action potential decreased drastically, whereas the maximum firing rate increased significantly (Supplementary Fig. 2c, d).

Synapse formation is an important step in the functional development of neurons in the brain. To assess synapses formed onto the excitatory neurons in ontogenetic radial clones, we examined the spontaneous miniature excitatory postsynaptic currents (mini-EPSCs; Supplementary Fig. 2e–k). Whereas the mean amplitude of mini-EPSCs detected at different developmental stages remained similar (Supplementary Fig. 2e, f), the frequency increased markedly as development progressed (Supplementary Fig. 2e, g). Moreover, the rise and decay of mini-EPSCs sped up significantly with development (Supplementary Fig. 2h–k). These results indicate progressive formation and maturation of synapses onto the excitatory neurons in ontogenetic radial clones during postnatal development.

Having found that the neurons in ontogenetic radial clones actively form synapses, we set out to determine whether sister neurons in individual radial clones form synapses with each other. Simultaneous whole-cell recordings were performed on two EGFP-expressing sister neurons in individual radial clones, the cell bodies of which were often more than 100  $\mu\text{m}$  apart (Fig. 2a). After the recordings were established, single as well as a train of action potentials were triggered in one of the neurons by current injection, whereas the other neuron was kept in voltage-clamp mode around  $-70\text{ mV}$  to record postsynaptic responses (Fig. 2b). If the sister neurons are synaptically connected, action potentials triggered in the presynaptic neuron should evoke synaptic currents in the postsynaptic neuron. Indeed, we found that action potentials triggered in one neuron (neuron 1) reliably evoked postsynaptic currents in the other neuron (neuron 2; Fig. 2a, b), indicating that these two sister neurons in a radial clone are connected. Moreover, we found that the connection between them was unidirectional, as action potentials triggered in neuron 2 failed to evoke detectable postsynaptic currents in neuron 1 (Fig. 2a, b). Similar results were obtained when the postsynaptic neuron was kept in current-clamp mode (Fig. 2c). The biophysical properties of these postsynaptic responses recorded between sister neurons in radial clones indicate that they are glutamate-receptor-mediated EPSCs. This was corroborated by pharmacological experiments using picrotoxin, D-(–)-2-amino-5-phosphonopentanoic acid (D-AP5) and 2,3-dioxo-6-nitro-1,2,3,4-tetrahydrobenzo[*f*]quinoxaline-7-sulphonamide (NBQX) (Fig. 2d)—the inhibitors of  $\gamma$ -aminobutyric acid (GABA)-A, *N*-methyl-D-aspartic acid (NMDA) and  $\alpha$ -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors, respectively, as well as carbenoxolone and



**Figure 2 | Synapse formation between sister excitatory neurons in ontogenetic radial clones.** **a**, Images of a pair of EGFP-expressing (green, middle) sister neurons in a radial clone in whole-cell configuration. Alexa-546-conjugated biocytin (red, right) was included in the recording pipette to confirm the cells being recorded and to reveal cell morphology. A differential interference contrast (DIC) image is shown to the left and arrowheads indicate two EGFP-expressing green sister neurons. Scale bars, 100  $\mu\text{m}$  and 50  $\mu\text{m}$  (left and middle, respectively). **b**, **c**, Sample traces of action potentials triggered in the presynaptic neurons (red) and EPSCs (**b**) or excitatory postsynaptic potentials (**c**) recorded in the postsynaptic neurons under voltage-clamp (**b**) or current-clamp (**c**) mode (green or black). The bold traces represent the average and the grey traces represent the individual recordings. Scale bars, 40 mV (red trace, vertical scale bar), 15 pA (green and black traces, vertical scale bar) and 200 ms (horizontal scale bar) (**b**), and 50 mV (trace 1, vertical scale bar), 1.5 mV (trace 2, vertical scale bar) and 50 ms (horizontal scale bar) (**c**). **d**, EPSC-blockage in postsynaptic neuron 2 by NBQX and D-AP5, but not by picrotoxin. Scale bar, 10 pA (vertical) and 5 ms (horizontal). **e**, Summary of the rate of connectivity between sister excitatory neuron pairs in individual radial clones.

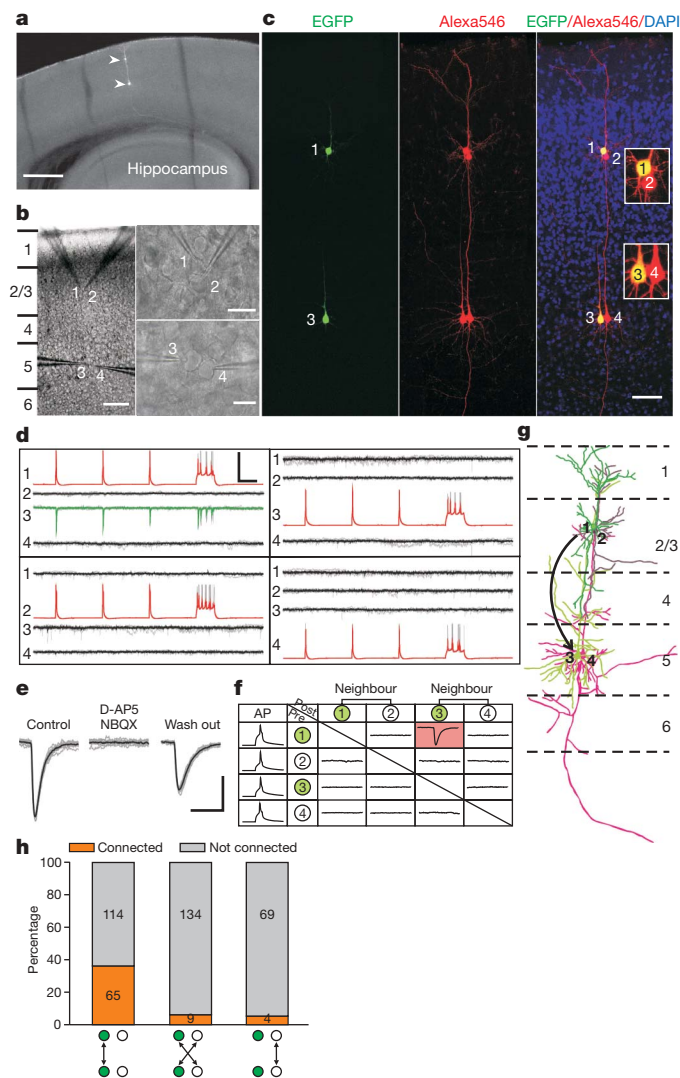
18 $\alpha$ -glycyrrhetic acid—two commonly used blockers of gap junctions (Supplementary Fig. 3).

Our results thus far indicate that sister excitatory neurons in individual radial clones in the developing neocortex develop unidirectional chemical synapses. We examined 101 pairs of sister neurons in individual radial clones at different developmental stages (Fig. 2e). Whereas the rate of finding a connected pair of sister neurons in a radial clone was rather low at the first postnatal (P) week (0 out of 34 at P1–P5, and 2 out of 21 at P6–P9), it increased markedly around the second postnatal week (7 out of 20 at P10–P13, and 9 out of 26 at P14–P17), suggesting that the second postnatal week is a critical period for the development of synaptic connections between sister neurons in a radial clone. This time window coincides with the critical period of functional circuit development in the neocortex<sup>24–26</sup>.

Previous studies suggested that the probability of detecting a synaptic connection between adjacent excitatory neurons in the mature neocortex is from 5% to 20% depending on the neocortical layer that they are located in and the neuronal subtype<sup>4–7,27</sup>. Sister excitatory neurons in a radial clone in the developing neocortex are located quite far away from each other (a hundred to a few hundred micrometres). Therefore, the probability of finding a synaptic connection between them would be far lower<sup>2</sup>. Nonetheless, we found that ~35% (16 out of 46) of sister excitatory neurons in a radial clone around P10 to P17 were connected, suggesting a propensity for radially aligned sister excitatory neurons to form synapses with each other.



To test this directly, we performed quadruple whole-cell recordings on two EGFP-expressing sister neurons in individual radial clones and two non-EGFP-expressing neurons adjacent to the EGFP-expressing sister neurons on the same side in the developing neocortex at P10–P17 (Fig. 3a–c and Supplementary Fig. 4). We targeted non-EGFP-expressing neighbouring neurons with radially aligned pyramid-shaped cell bodies because they were most likely to be excitatory neurons and



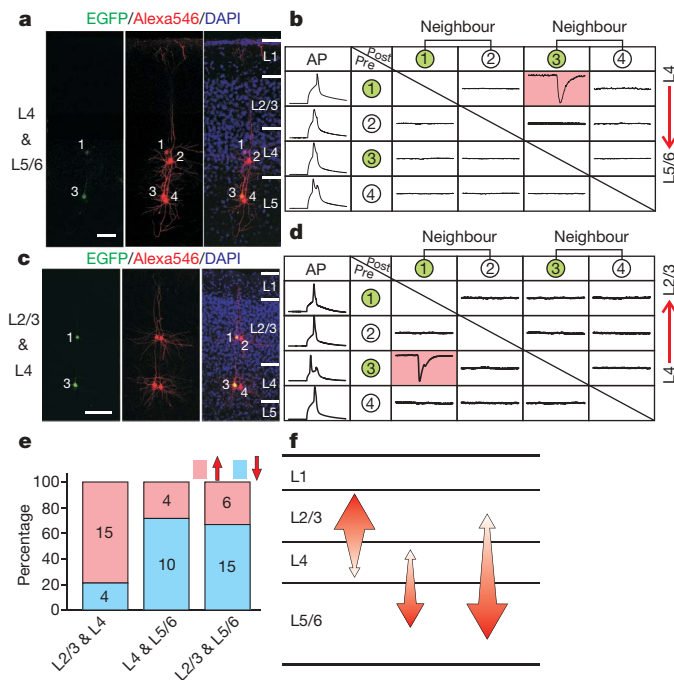
**Figure 3 | A strong preference for synapse formation between sister excitatory neurons in ontogenetic radial clones.** **a**, Image of two sister neurons in an ontogenetic radial clone in the developing neocortex. Scale bar, 1 mm. **b**, **c**, DIC (**b**) and fluorescence (**c**) images of a quadruple whole-cell recording of two EGFP-expressing sister neurons (1 and 3) shown in **a** and two non-EGFP-expressing neurons (2 and 4) adjacent to the sisters at the same side. Scale bars, 200  $\mu$ m (left) and 20  $\mu$ m (right) (**b**) and 100  $\mu$ m (**c**). **d**, Sample traces of action potentials (red) triggered in the presynaptic neurons and EPSCs (green and black) recorded in the postsynaptic neurons. Scale bars, 80 mV, 20 pA and 200 ms. **e**, Blockage of EPSCs by NBQX and D-AP5. Scale bars, 10 pA and 5 ms. **f**, Summary of the synaptic connection detected in this quadruple recording. Green circles indicate EGFP-expressing sister excitatory neurons, white circles indicate non-EGFP-expressing neighbouring excitatory neurons. The average traces of the postsynaptic responses are shown in the rectangle. The red background indicates detection of the EPSC. Sample traces of action potentials (AP) are shown to the left. 'Pre' and 'post' represent presynaptic and postsynaptic neurons, respectively. **g**, Reconstruction of the neurons in this quadruple recording. The arrow indicates the detection of the synaptic connection among the four neurons. **h**, The rate of synaptic connectivity observed between sister excitatory neurons in ontogenetic radial clones and between radially situated non-sister excitatory neurons.

the progeny of different progenitor cells. Furthermore, the morphology and biophysical properties of these control neurons confirmed that they were excitatory neurons (Fig. 3d–g). Hence they served as adjacent non-sibling controls. Once all four recordings were established, action potentials were sequentially triggered in one of the four neurons and the postsynaptic responses were then measured in the other three neurons to probe synapses formed among them. As shown in Fig. 3d, e, when action potentials were triggered in EGFP-expressing neuron 1, glutamate-receptor-mediated EPSCs were recorded only in its sister neuron 3 (Fig. 3d, e). Despite the nearly complete overlap of the dendritic arbours of adjacent neurons 3 and 4 (Fig. 3g), no EPSC was detected in the non-sibling neuron 4 (Fig. 3d). Furthermore, action potentials triggered in all three remaining neurons (neuron 2, 3 and 4) failed to evoke any detectable EPSCs in the other three neurons (Fig. 3d). These results indicate that a unidirectional synaptic connection selectively develops between two sister excitatory neurons, but not between non-sisters, in the developing neocortex (Fig. 3g).

We analysed a total of 179 pairs of radially aligned EGFP-expressing sister excitatory neurons and their neighbouring non-sibling neurons (Fig. 3h). Among them, 36.9% (65 out of 179) of sister neurons in a radial clone were connected. In contrast, only 6.3% (9 out of 143) of radially situated non-sister excitatory neurons (one EGFP-expressing and one non-EGFP-expressing) were connected (Fig. 3h). These results clearly demonstrate that sister excitatory neurons in a radial clone have a strong preference to form synapses with each other instead of with adjacent non-sister neurons, and the rate of connectivity between distant, radially situated non-sister excitatory neurons is low. Consistent with this low rate of connectivity, only 5.5% (4 out of 73) of any two randomly selected radially situated non-EGFP-expressing neurons over at least 100  $\mu$ m apart were connected (Fig. 3h).

Previous studies have shown the overall organization of the excitatory neuron microcircuit in the mature neocortex<sup>28,29</sup>. Thalamic input enters primarily into layer 4 (the first station of sensory processing). Layer 4 excitatory neurons send ascending projections to pyramidal neurons in layer 2/3 (the second station of columnar processing), which provide a prevalent descending projection to layer 5/6 pyramidal neurons (the third station of columnar processing). Descending and ascending excitatory connections also exist between layer 4 and layer 5/6. In addition to the sister neurons radially situated in layer 2/3 and layer 5/6 (Fig. 3a–g), our data set of quadruple recordings contained radially aligned sister neurons located in layer 4 and layer 5/6 (Fig. 4a, b and Supplementary Fig. 5a) as well as in layer 2/3 and layer 4 (Fig. 4c, d and Supplementary Fig. 5b). These experiments allowed us to address the interlaminar direction preference of synaptic connectivity formed within ontogenetic radial clones of excitatory neurons. We found that 15 out of 21 connected sister excitatory neuron pairs located in layer 2/3 and layer 5/6, and 10 out of 14 pairs of those located in layer 4 and layer 5/6 formed synapses in the descending direction (that is, from layer 2/3 to layer 5/6 and from layer 4 to layer 5/6), whereas 15 out of 19 pairs of connected sister excitatory neurons located in layer 2/3 and layer 4 formed synapses in the ascending direction (that is, from layer 4 to layer 2/3; Fig. 4e, f). These results indicate that the synaptic connection formed among sister excitatory neurons in ontogenetic radial clones is rather specific. Moreover, these results demonstrate that the specificity of synaptic connectivity formed within ontogenetic radial clones of excitatory neurons in the developing neocortex is highly similar to that in the mature neocortex.

The concept of the column has cast a dominant influence on our understanding of the functional organization of the neocortex<sup>1,2</sup>. From its inception, the concept of the functional column has been considered on both a macroscopic and microscopic scale. However, most of our knowledge about functional columns and neocortical maps comes from measurements with limited spatial resolution. Recent *in vivo* Ca<sup>2+</sup> imaging studies elegantly demonstrated that even adjacent neurons can have distinct physiological properties<sup>9–11</sup>, indicating that neocortical maps are built with single-neuron precision. In this study, we found that sister excitatory neurons in individual radial clones in



**Figure 4 | A highly specific microcircuit forms among sister excitatory neurons in ontogenetic radial clones.** **a–d**, Quadruple recording of sister excitatory neurons in individual ontogenetic radial clones located in layer (L) 4 and layer 5/6 (**a**, **b**) or layer 2/3 and layer 4 (**c**, **d**), and their adjacent non-sister excitatory neurons. See Fig. 3 legend for details. **e**, **f**, Summary of the direction of synaptic connections observed between sister excitatory neurons in individual ontogenetic radial clones. The size of the arrows in **f** reflects the abundance of the connection.

the developing mouse neocortex preferentially develop highly specific synaptic connections with each other, creating radial columnar micro-architectures of interconnected neuron ensembles with single-neuron resolution. The high degree of similarity in the direction of interlaminar connectivity between the synapses formed within individual ontogenetic radial clones and those observed in the mature neocortex suggests that these radial clones contribute to the formation of precise functional columnar architectures in the neocortex. Along this line, variations in the neurogenesis and emergence of ontogenetic radial clones during early neocortical development<sup>30</sup> may underlie differences that have been observed in the functional organization of the mature neocortex across species<sup>10</sup>.

## METHODS SUMMARY

Replication-incompetent EGFP-expressing retroviruses (a kind gift from S. C. Noctor and F. H. Gage) were intraventricularly injected into E12–E13 mouse embryos as previously described<sup>13</sup>. Acute cortical slices were prepared at different postnatal stages. Multiple-electrode whole-cell recordings were performed on EGFP-expressing neurons in individual radial clones and their near neighbours under visual guidance of epi-fluorescence and infrared differential interference contrast (IR-DIC) illumination. Recordings were collected and analysed using Axopatch 700B amplifier and pCLAMP10 software (Molecular Devices) and Igor 5 software (Wavemetrics Inc.). The morphology of neurons was reconstructed with confocal laser scanning microscopy using FluoView (Olympus), Neurolucida (MicroBrightField Inc.) and Photoshop (Adobe Systems). Data were presented as mean and s.e.m. and statistical differences were determined using the nonparametric Mann–Whitney–Wilcoxon test.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 22 October; accepted 11 December 2008.  
Published online 8 February 2009.

- Hubel, D. H. & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol. (Lond.)* **160**, 106–154 (1962).
- Mountcastle, V. B., Davies, P. W. & Berman, A. L. Response properties of neurons of cat's somatic sensory cortex to peripheral stimuli. *J. Neurophysiol.* **20**, 374–407 (1957).

- Kozloski, J., Hamzei-Sichani, F. & Yuste, R. Stereotyped position of local synaptic targets in neocortex. *Science* **293**, 868–872 (2001).
- Markram, H., Lubke, J., Frotscher, M., Roth, A. & Sakmann, B. Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex. *J. Physiol. (Lond.)* **500**, 409–440 (1997).
- Song, S., Sjostrom, P. J., Reigl, M., Nelson, S. & Chklovskii, D. B. Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biol.* **3**, e68 (2005).
- Thomson, A. M., West, D. C., Wang, Y. & Bannister, A. P. Synaptic connections and small circuits involving excitatory and inhibitory neurons in layers 2–5 of adult rat and cat neocortex: triple intracellular recordings and biocytin labelling *in vitro*. *Cereb. Cortex* **12**, 936–953 (2002).
- Yoshimura, Y., Dantzker, J. L. & Callaway, E. M. Excitatory cortical neurons form fine-scale functional networks. *Nature* **433**, 868–873 (2005).
- Yoshimura, Y. & Callaway, E. M. Fine-scale specificity of cortical networks depends on inhibitory cell type and connectivity. *Nature Neurosci.* **8**, 1552–1559 (2005).
- Ohki, K. *et al.* Highly ordered arrangement of single neurons in orientation pinwheels. *Nature* **442**, 925–928 (2006).
- Ohki, K., Chung, S., Ch'ng, Y. H., Kara, P. & Reid, R. C. Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. *Nature* **433**, 597–603 (2005).
- Sato, T. R., Gray, N. W., Mainen, Z. F. & Svoboda, K. The functional microarchitecture of the mouse barrel cortex. *PLoS Biol.* **5**, e189 (2007).
- Maldonado, P. E., Godecke, I., Gray, C. M. & Bonhoeffer, T. Orientation selectivity in pinwheel centers in cat striate cortex. *Science* **276**, 1551–1555 (1997).
- Noctor, S. C., Flint, A. C., Weissman, T. A., Dammerman, R. S. & Kriegstein, A. R. Neurons derived from radial glial cells establish radial units in neocortex. *Nature* **409**, 714–720 (2001).
- Noctor, S. C., Martinez-Cerdeno, V., Ivic, L. & Kriegstein, A. R. Cortical neurons arise in symmetric and asymmetric division zones and migrate through specific phases. *Nature Neurosci.* **7**, 136–144 (2004).
- Miyata, T., Kawaguchi, A., Okano, H. & Ogawa, M. Asymmetric inheritance of radial glial fibers by cortical neurons. *Neuron* **31**, 727–741 (2001).
- Malatesta, P., Hartfuss, E. & Gotz, M. Isolation of radial glial cells by fluorescent-activated cell sorting reveals a neuronal lineage. *Development* **127**, 5253–5263 (2000).
- Tamamaki, N., Nakamura, K., Okamoto, K. & Kaneko, T. Radial glia is a progenitor of neocortical neurons in the developing cerebral cortex. *Neurosci. Res.* **41**, 51–60 (2001).
- Rakic, P. Mode of cell migration to the superficial layers of fetal monkey neocortex. *J. Comp. Neurol.* **145**, 61–83 (1972).
- Rakic, P. Specification of cerebral cortical areas. *Science* **241**, 170–176 (1988).
- Kornack, D. R. & Rakic, P. Radial and horizontal deployment of clonally related cells in the primate neocortex: relationship to distinct mitotic lineages. *Neuron* **15**, 311–321 (1995).
- Cepko, C. L. *et al.* Studies of cortical development using retrovirus vectors. *Cold Spring Harb. Symp. Quant. Biol.* **55**, 265–278 (1990).
- Walsh, C. & Cepko, C. L. Clonally related cortical cells show several migration patterns. *Science* **241**, 1342–1345 (1988).
- Walsh, C. & Cepko, C. L. Clonal dispersion in proliferative layers of developing cerebral cortex. *Nature* **362**, 632–635 (1993).
- Hensch, T. K. Critical period plasticity in local cortical circuits. *Nature Rev. Neurosci.* **6**, 877–888 (2005).
- Micheva, K. D. & Beaulieu, C. Quantitative aspects of synaptogenesis in the rat barrel field cortex with special reference to GABA circuitry. *J. Comp. Neurol.* **373**, 340–354 (1996).
- Stern, E. A., Maravall, M. & Svoboda, K. Rapid development and plasticity of layer 2/3 maps in rat barrel cortex *in vivo*. *Neuron* **31**, 305–315 (2001).
- Sjostrom, P. J., Turrigiano, G. G. & Nelson, S. B. Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron* **32**, 1149–1164 (2001).
- Thomson, A. M. & Bannister, A. P. Interlaminar connections in the neocortex. *Cereb. Cortex* **13**, 5–14 (2003).
- Douglas, R. J. & Martin, K. A. Neuronal circuits of the neocortex. *Annu. Rev. Neurosci.* **27**, 419–451 (2004).
- Kriegstein, A., Noctor, S. & Martinez-Cerdeno, V. Patterns of neural stem and progenitor cell division may underlie evolutionary cortical expansion. *Nature Rev. Neurosci.* **7**, 883–890 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank A. L. Joyner, J. Kaltschmidt, Y. Hayashi and Y. Chin for comments on the manuscript; and S. C. Noctor and F. H. Gage for providing the 293gp NiT–GFP retrovirus packaging cell line; and the Shi laboratory members for insightful discussion. We are grateful for support from the March of Dimes Foundation, the Whitehall Foundation, the Klingenstein Foundation, the DANA Foundation, the Autism Speaks Foundation, the National Alliance for Research on Schizophrenia and Depression (NARSAD) and the National Institutes of Health (to S.-H.S.).

**Author Contributions** Y.-C.Y. and S.-H.S. conceived the experiments. Y.-C.Y. conducted the electrophysiology and imaging experiments. R.S.B. and X.W. helped with *in utero* virus injection. Y.-C.Y. and S.-H.S. analysed the data and wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to S.-H.S. ([shis@mskcc.org](mailto:shis@mskcc.org)).

## METHODS

**Retroviral infection.** Replication-incompetent EGFP-expressing retrovirus was produced from a stably transfected packaging cell line (293gp NIT-GFP; a kind gift from S. C. Noctor and F. H. Gage) as previously published<sup>15</sup>. Animals were maintained according to protocols approved by the Institutional Animal Care and Use Committee at the Memorial Sloan Kettering Cancer Centre. Uterine horns of E12–E13 gestation stage pregnant CD-1 mice (Charles River Laboratories) were exposed in a clean environment. Retrovirus ( $\sim 1.0 \mu\text{l}$ ) with Fast green ( $2.5 \text{ mg ml}^{-1}$ , Sigma) was injected into the embryonic cerebral ventricle through a bevelled, calibrated glass micropipette (Drummond Scientific). After injection, the peritoneal cavity was lavaged with  $\sim 10 \text{ ml}$  warm PBS (pH 7.4) containing antibiotics, the uterine horns were replaced, and the wound was closed.

**Electrophysiology.** Embryos that received retroviral injections were delivered naturally. Brains were removed at different postnatal days and acute cortical slices were prepared at  $\sim 300 \mu\text{m}$  in artificial cerebral spinal fluid containing (in mM): 126 NaCl, 3 KCl, 1.25  $\text{KH}_2\text{PO}_4$ , 1.3  $\text{MgSO}_4$ , 3.2  $\text{CaCl}_2$ , 26  $\text{NaHCO}_3$  and 10 glucose, bubbled with 95%  $\text{O}_2$ , 5%  $\text{CO}_2$  on a Vibratome at  $4^\circ\text{C}$  (Leica Microsystems). Slices were first recovered in an interface chamber at  $35^\circ\text{C}$  for at least 1 h and then kept at room temperature before being transferred to a recording chamber containing artificial cerebral spinal fluid at  $34^\circ\text{C}$ . An infrared-DIC microscope (Olympus) equipped with epifluorescence illumination, a charge-coupled device camera, and two water immersion lenses ( $\times 10$  and  $\times 60$ ) were used to visualize and target recording electrodes to EGFP-expressing sister cells in radial clones and their nearby control cells. Glass recording electrodes ( $7\text{--}9 \text{ M}\Omega$  resistance) were filled with an intracellular solution consisting of 130 mM potassium-gluconate, 6 mM KCl, 2 mM  $\text{MgCl}_2$ , 0.2 mM EGTA, 10 mM HEPES, 2.5 mM  $\text{Na}_2\text{ATP}$ , 0.5 mM  $\text{Na}_2\text{GTP}$ , 10 mM potassium-phosphocreatine and 0.5% Alexa-546-conjugated biocytin (Invitrogen) (pH 7.25 and  $295 \text{ mOsm kg}^{-1}$ ). All recordings had access resistance less than  $30 \text{ M}\Omega$ . In all dual, triple and quadruple recordings, connections between neuron pairs were assessed by injecting current to evoke action potentials in one of the cells kept in current-clamp while testing for postsynaptic responses in other

cells during voltage-clamp recording at  $-70 \text{ mV}$  unless specified. For every possible pair, connections were tested in both directions for at least 20 trials, with both single and trains of action potential being generated in each presynaptic neuron. In some experiments, picrotoxin ( $50 \mu\text{M}$ ), D-AP5 ( $50 \mu\text{M}$ ) and NBQX ( $10 \mu\text{M}$ ) (Tocris Biosciences) were added to the bath to block GABA-A, NMDA and AMPA receptors, respectively. Carbenoxolone ( $100 \mu\text{M}$ ) and  $18\alpha$ -glycyrrhetinic ( $25 \mu\text{M}$ ) (Sigma) were used to block gap junctions. Recordings were collected and analysed using Axopatch 700B amplifier and pCLAMP10 software (Molecular Devices) and Igor 5 software (Wavemetrics Inc.). Spontaneous mini-EPSCs were analysed using mini Analysis Program (Synaptosoft Inc.). At all developmental stages, the EPSC decay was best described by the sum of two exponential functions:

$$A(t) = A_{\text{fast}}(\exp(-t/\tau_{\text{fast}})) + A_{\text{slow}}(\exp(-t/\tau_{\text{slow}}))$$

in which  $\tau_{\text{fast}}$  and  $\tau_{\text{slow}}$  are the decay time constant of the fast and slow components,  $A_{\text{fast}}$  and  $A_{\text{slow}}$  are their respective amplitudes, and  $t$  represents time. Data were presented as mean and s.e.m. and statistical differences were determined using nonparametric Mann–Whitney–Wilcoxon test.

**Confocal microscopy.** For morphological analysis of radial clones, animals were perfused with cold PBS (pH 7.4) followed by 4% paraformaldehyde in PBS. Brains were recovered and sectioned using a Vibratome (Leica Microsystems). Sections were blocked in 10% serum and 0.1% Triton-X in PBS, and incubated with a chicken anti-GFP antibody (Aves Lab Inc.) overnight at  $4^\circ\text{C}$ . An Alexa-488-conjugated secondary antibody (Invitrogen) was then used to visualize GFP for high resolution morphological analysis. In whole-cell patch clamp recording experiments Alexa-546-conjugated biocytin was loaded onto neurons with a whole cell recording pipette, slices were fixed in 4% paraformaldehyde in PBS (pH 7.4) and biocytin was visualized with Cy3-conjugated streptavidin (Invitrogen). Z-series images were taken at  $1 \mu\text{m}$  using an Olympus FV1000 confocal laser scanning microscope. Images were analysed using FluoView (Olympus), Neurolucida (MicroBrightField, Inc.) and Photoshop (Adobe Systems).



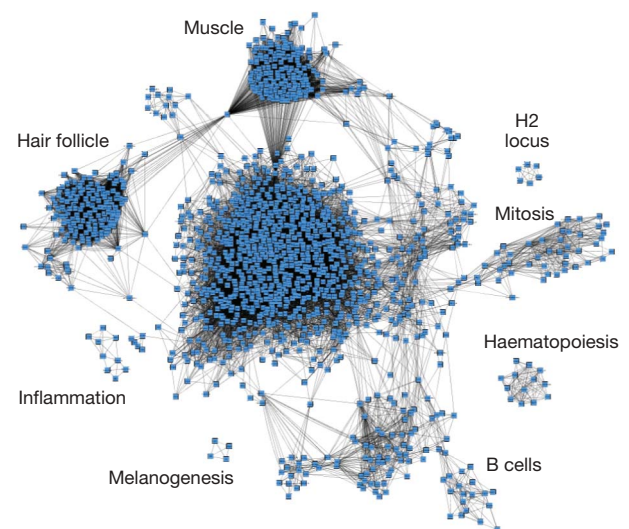
# Genetic architecture of mouse skin inflammation and tumour susceptibility

David A. Quigley<sup>1</sup>, Minh D. To<sup>1,2</sup>, Jesús Pérez-Losada<sup>3</sup>, Facundo G. Pelorosso<sup>1</sup>, Jian-Hua Mao<sup>1,4</sup>, Hiroki Nagase<sup>5,6</sup>, David G. Ginzing<sup>1</sup> & Allan Balmain<sup>1</sup>

Germline polymorphisms in model organisms and humans influence susceptibility to complex trait diseases such as inflammation and cancer<sup>1–4</sup>. Mice of the *Mus spretus* species are resistant to tumour development, and crosses between *M. spretus* and susceptible *Mus musculus* strains have been used to map locations of genetic variants that contribute to skin cancer susceptibility<sup>4–6</sup>. We have integrated germline polymorphisms with gene expression in normal skin from a *M. musculus* × *M. spretus* backcross to generate a network view of the gene expression architecture of mouse skin. Here we demonstrate how this approach identifies expression motifs that contribute to tissue organization and biological functions related to inflammation, haematopoiesis, cell cycle control and tumour susceptibility. Motifs associated with inflammation, epidermal barrier function and proliferation are differentially regulated in backcross mice susceptible or resistant to tumour development. The intestinal stem cell marker *Lgr5* is identified as a candidate master regulator of the hair follicle, and the vitamin D receptor (*Vdr*) is linked to coordinated control of epidermal barrier function, inflammation and tumour susceptibility.

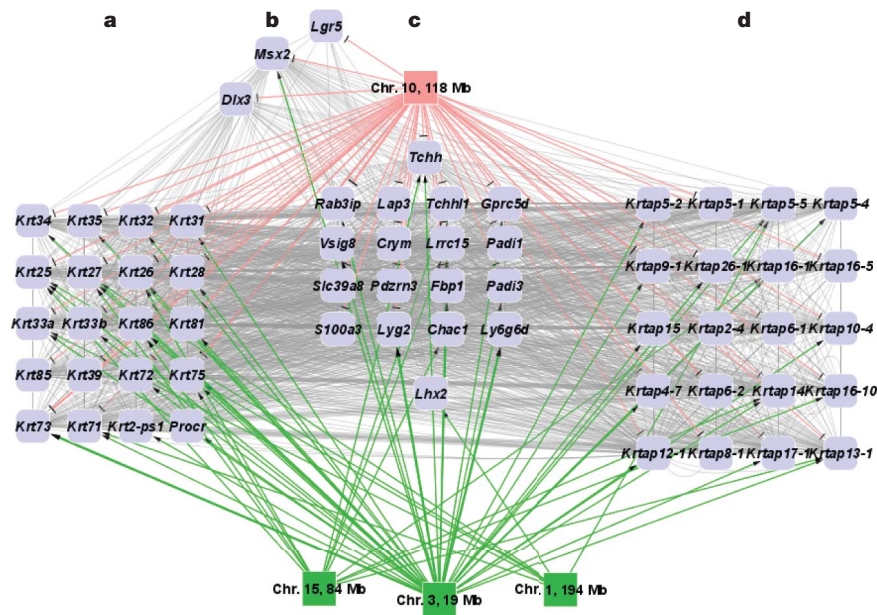
Skin tumours were induced on the dorsal back skin of 71 *M. spretus* × *M. musculus* backcross mice ((SPRET/Ei × FVB/N) × FVB/N, hereafter referred to as FVBBX mice, see Methods Summary). We performed a combined correlation and linkage analysis of messenger RNA expression in uninvolved tail skin to characterize the genetic architecture of gene expression regulation in FVBBX mice. This approach identifies genetic loci that influence gene expression (eQTL), with either *cis*- or *trans*-acting effects. Several loci were hot-spots for significant *trans*-eQTL, including chromosome 2 at 85 Mb, chromosome 1 at 29 Mb, and chromosome 15 at 73 Mb (Supplementary Fig. 1). However, not all transcripts influenced by a given locus are functionally related to each other or co-regulated. We used a network analysis approach to identify loci that control functionally related groups of genes. We started by representing FVBBX tail skin RNA expression as a network where significantly correlated transcripts are drawn as nodes connected by an edge (see Methods Summary). We then identified fully connected gene sets (cliques) that were enriched for functions of interest (Fig. 1). These sub-networks were highly enriched for genes representing either structural or functional components of the skin. These included skin-resident cell types (hair follicles, muscle, melanocytes, haematopoietic cells) and physiological (for example, inflammation) or cellular (for example, cell cycle) functions (Supplementary Fig. 4 and Supplementary Table 3). We next identified the genetic loci associated with control of these different expression motifs (see Methods Summary).

A network comprising 62 genes, most of which are involved in hair follicle biology, was identified (Fig. 2 and Supplementary Table 4). Numerous Type I and Type II hair follicle keratin genes such as *Krt32* and *Krt73* were present (Fig. 2a) along with genes involved in keratin structure and differentiation (Fig. 2c) and many keratin-associated proteins (Fig. 2d). Also identified were *Msx2*, *Dlx3* and *Lhx2* (Fig. 2b); these homeobox genes have been implicated in transcriptional control of hair follicle morphogenesis. Notably, epidermal keratins such as *Krt5* and *Krt14* are under separate genetic control and their expression was not correlated with hair follicle keratins. We extracted epithelial tissue from the tails of *M. spretus*, *M. musculus* and SPRET/Ei × FVB/N mice (hereafter referred to as SPRET/Ei, FVB/N and F<sub>1</sub>, respectively) to investigate the hypothesis that expression of hair follicle keratins is under complex genetic control. Expression of Type I and Type II hair follicle keratins was significantly different in these parental mice (*q*-values = 0, see Methods Summary); expression was approximately eightfold higher in FVB/N mice than in SPRET/Ei. Expression of these genes was significantly higher in F<sub>1</sub> mice than in either of the parental strains (Supplementary Fig. 2a),



**Figure 1 | A visual representation of the FVBBX gene expression network.** Network edges (black lines) connect significantly correlated genes (blue squares) that are co-regulated in cliques with five or more members. Labels identify clusters significantly enriched for functional or structural properties. Thousands of genes have significant eQTL, which are not shown here to preserve clarity.

<sup>1</sup>Helen Diller Family Comprehensive Cancer Center, <sup>2</sup>Department of Surgery, University of California San Francisco, San Francisco, California 94115, USA. <sup>3</sup>Departamento de Medicina, Facultad de Medicina, Universidad de Salamanca, Campus Miguel de Unamuno s/n, 37007 Salamanca, Spain. <sup>4</sup>Life Sciences Division, Lawrence Berkeley National Laboratory, University of California, Berkeley, California 94720, USA. <sup>5</sup>Advanced Research Institute for the Sciences and Humanities, Nihon University, Tokyo, Japan. <sup>6</sup>Department of Cancer Genetics, Roswell Park Cancer Institute, Buffalo, New York 14263, USA.



**Figure 2 | The hair follicle gene expression and linkage network.** Network edges (black lines) connect significantly correlated genes (blue squares) from several overlapping highly correlated cliques. Red and green edges connect eQTL loci (red and green nodes) with significant linkage to genes. Green arrow-terminated edges indicate the gene is upregulated by the SPRET/Ei

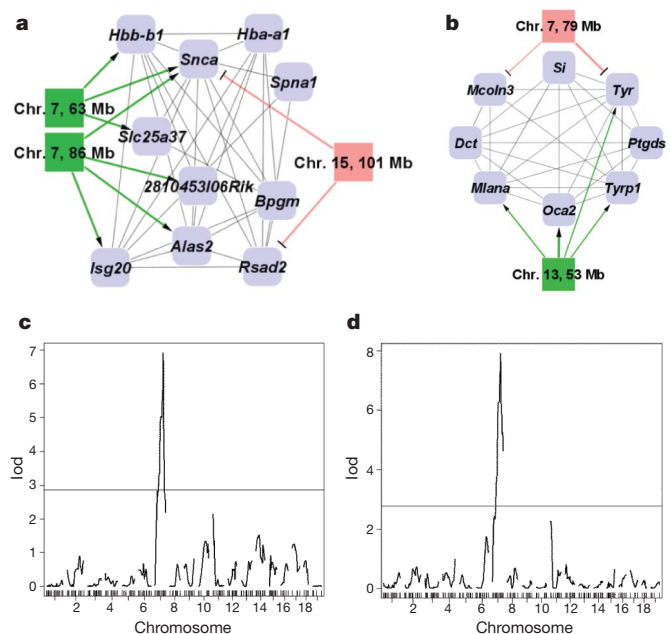
allele; red bar-terminated edges indicate the SPRET/Ei downregulates the gene. See also Supplementary Table 4. **a**, Type I and Type II keratins. **b**, Transcription factors *Msx2* and *Dlx3* and stem cell marker *Lgr5*. **c**, Genes associated with keratin biology and other genes. **d**, Keratin-associated proteins.

indicating that expression of these structural protein genes is under complex genetic control. The transcription factors *Dlx3* and *Msx2* were also significantly differentially expressed in the same directions.

The eQTLs in this hair follicle network controlling the largest number of genes were located on chromosome 10 at 118 Mb and on chromosome 3 at 19 Mb. Inheritance of the SPRET/Ei allele on chromosome 3 at 19 Mb is associated with higher expression levels of many genes in this network, whereas the SPRET/Ei allele on the chromosome 10 locus decreases expression of these genes. *Lgr5* (also known as *Gpr49*), a member of this network, is physically located on chromosome 10 at 115 Mb and has a strong *cis*-eQTL on chromosome 10 at 118 Mb ( $-\log_{10}P = 4.3$ ,  $q$ -value  $< 0.05$ , Fig. 2b). *Lgr5* is a G-protein-coupled receptor recently identified as a stem cell marker in colon and small intestine<sup>7</sup>. *Lgr5* is the only gene with a significant *cis*-eQTL in this sub-network (the other eQTL effects act in *trans*) and *Lgr5* is plausibly the causative polymorphic gene on chromosome 10 that influences expression of these hair follicle genes. *Lgr5* microarray results were validated by a Taqman assay, and *cis*-regulation of *Lgr5* expression in F<sub>1</sub> mice was verified by a *cis*-*trans* test (Supplementary Figs 5 and 6). Sequencing of *Lgr5* coding exons in SPRET/Ei and FVB/N identified four non-conservative differences between these two species (Supplementary Table 5). In agreement with this network view of the hair follicle, *Lgr5* was recently identified as a hair follicle stem cell marker by lineage tracing<sup>8</sup>. Under normal homeostatic conditions, *Lgr5*-positive skin cells give rise to lineages of the hair follicle, but not to sebaceous glands or interfollicular epidermis<sup>8</sup>.

Our approach also identified motifs associated with genetic control of haematopoietic cells and melanocytes (Fig. 3a, b and Supplementary Tables 6 and 7). A haemoglobin production network including the haemoglobin isoform *Hbb-b1*,  $\alpha$ -synuclein (*Snca*) and aminolevulinic acid synthase 2 (*Alas2*) was identified with eQTL that peak on chromosome 7 at 63 Mb and 86 Mb. The observation of eQTL for expression of red-blood-cell-specific genes in the skin indicated that specific alleles control the number of skin-resident blood cells. This could be due to a physiological process that results in a specific homing of haematopoietic cells to the skin of particular

animals, or may reflect increased systemic red blood cell production that is under genetic control. To test these alternative possibilities we generated a second FVBBX population ( $n = 83$ ) and performed QTL analysis on complete blood count (CBC) parameters (see Methods Summary). Mean corpuscular volume showed a highly significant peak logarithm of odds (LOD) score of 6.8 on chromosome 7 at 79 Mb and mean corpuscular haemoglobin content had a peak



**Figure 3 | Haematopoiesis gene expression and linkage networks are confirmed by QTL results.** **a**, **b**, Haematopoietic (**a**) and melanosomal (**b**) eQTL networks in FVBBX epidermal RNA, drawn as in Fig. 2. See also Supplementary Tables 6 and 7. **c**, **d**, Genome-wide plot of LOD scores from an 83-mouse FVBBX cohort for mean corpuscular volume (**c**) and MCHC (**d**), where the horizontal line represents genome-wide significance.



LOD score of 8.2 on chromosome 7 at 86 Mb (corrected  $P$  values  $< 0.001$ , Fig. 3c, d). This supports the hypothesis that genetic factors, rather than environmental factors such as skin inflammation, control blood cell phenotypes in the circulation and within the skin. It is also likely that variation in the numbers of melanocytes in the skin is under the control of genetic loci on chromosome 7 (Fig. 3b and Supplementary Table 7).

We next identified networks associated with susceptibility to the development of skin tumours. The overall dorsal papilloma count in FVB/BX mice after 20 weeks of dimethyl benzantracene (DMBA) and tetradecanoyl-phorbol acetate (TPA) treatment ranged from 0 to 35 (mean = 5.4, s.d. = 6.7). We assigned mice with zero papillomas at 20 weeks to a low-susceptibility group and mice with eight or more papillomas to a high-susceptibility group and performed differential expression analysis of tail epidermal RNA. A total of 371 putative susceptibility genes were identified at a false discovery rate (FDR) cutoff of 10%. This list was significantly enriched for epidermal keratinization and barrier function, the mitotic cycle and DNA replication, and interleukin 1 antagonist activity (Supplementary Table 8). Many genes involved in mitosis and cell proliferation (including *Aurkb*, *Bub1*, *Nuf2*, *Cdca8*, *Ect2*) were significantly co-regulated and associated with susceptibility, but do not share a statistically significant eQTL in normal skin. This motif includes *Aurora-A* (also known as *Aurka*), which has previously been identified as a genetic modifier of papilloma number<sup>5</sup>.

We identified a locus on distal chromosome 15 at 101 Mb that influences expression of many of these genes (Fig. 4, Supplementary Table 9). This susceptibility network includes *Il1f6* and *Il1f5*, ligands for the interleukin 1 receptor that can act positively or negatively in the inflammatory response, depending on the context<sup>9</sup>. The interleukin 1 receptor is the major mediator of inflammation in the skin<sup>10</sup>. Mice with a SPRET/Ei allele at this locus had higher expression of inflammation antagonists *Il1f5* and *Il1f6* and lower expression of pro-inflammatory genes such as *Pde4b* (ref. 11). This pattern is compatible with the overall resistance of *M. spretus* to inflammation noted by others<sup>12</sup>. We observed that higher expression of inflammation antagonists and lower expression of *Pde4b* was

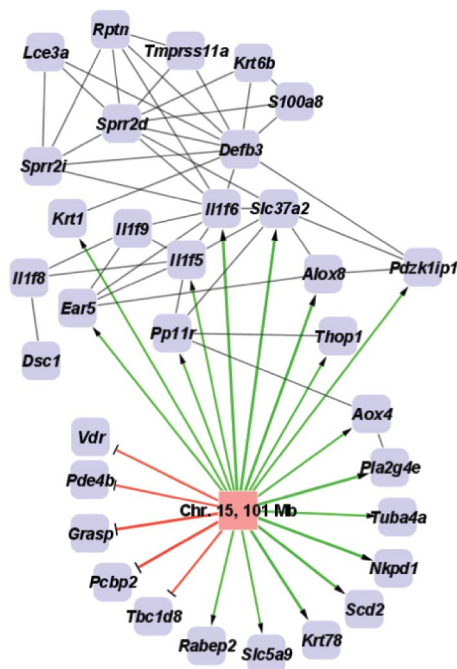
associated with high papilloma count, indicating that in this context the *M. spretus* allele increases susceptibility, counter to the idea that inflammation is invariably associated with increased tumour susceptibility<sup>13,14</sup>. However, anti-inflammatory drugs can have contradictory effects on skin tumour development<sup>15</sup> and overexpression of pro-inflammatory cytokines such as *Il1* can prevent skin tumour formation in mouse models of chemically induced skin cancer<sup>10</sup>. The role of inflammation in skin cancer is therefore highly complex, with possibly different consequences associated with acute or chronic inflammatory conditions.

Several genes related to skin barrier function were highly correlated with the inflammation network, although genetic linkage to chromosome 15 at 101 Mb was only detected at a lower level of stringency than that used for this analysis. These included *Rptn* and small proline-rich proteins *Sprr2d* and *Sprr2i*, major constituents of the epidermal barrier and cornified cell envelope. Disruption of skin barrier function leads to inflammation in epithelial tissue<sup>16–18</sup>, but the exact mechanisms linking these processes are unclear. This network also includes the microbial peptide gene defensin beta 3 (*Defb3*), implicated in inflammation and psoriasis<sup>19</sup>, and the anti-microbial calcium-binding gene *S100a8*, which complexes with *S100a9* to form a pro-inflammatory, anti-bacterial cytokine present in keratinocytes that are inflamed or infected<sup>20</sup>.

The best candidate master regulator of this network is the gene encoding the vitamin D receptor (*Vdr*). This gene is physically located on chromosome 15 at 97.7 Mb and shows a strong *cis*-eQTL ( $-\log_{10}P = 5.8$ ,  $q$ -value  $< 0.001$ ). Sequencing *Vdr* identified several non-conservative sequence changes between SPRET/Ei and FVB/N in a phosphorylation domain and the hormone-binding domain (Supplementary Figs 7 and 8). *Vdr* microarray results were validated by a Taqman assay, and *cis*-regulation of *Vdr* in F<sub>1</sub> mice was verified by a *cis*-*trans* test (Supplementary Figs 5 and 6). Low levels of *Vdr* in FVB/BX mice were associated with higher tumour susceptibility. Epidemiological studies have demonstrated an association between low serum levels of vitamin D and human cancer susceptibility, emphasizing an important role for vitamin D in human cancer prevention<sup>21</sup>. The vitamin D signalling pathway has been linked directly to control of skin barrier function<sup>22</sup>, inflammation and microbial defence<sup>23</sup>, and pathways involved in stem cell growth<sup>24</sup>. Importantly, loss of the *Vdr* gene in knockout mice causes the phenotype that would be predicted from the network shown in Fig. 4, namely increased skin tumour susceptibility<sup>25</sup>. This indicates that *Vdr* is a master regulator of tissue damage and acute inflammation. These results emphasize the value of integrated systems analysis of polymorphisms and gene expression to identify groups of interacting genes and candidate master regulators that contribute to individual cancer susceptibility. Applying this approach to human samples may reveal the major genetic components of cancer susceptibility that remain to be discovered.

## METHODS SUMMARY

Male SPRET/Ei and female FVB/N mice from the Jackson Laboratory were crossed. Female F<sub>1</sub> hybrids were mated to male FVB/N mice. Seventy-one back-cross mice (8–12 weeks old) received a dose of DMBA (25 µg per mouse in 200 µl acetone applied to shaved dorsal back skin). One week after initiation, tumours were promoted with TPA (200 µl of  $10^{-4}$  M solution in acetone) twice weekly for 20 weeks. Susceptibility groups were defined by papilloma count at 20 weeks (low: zero at 20 weeks,  $n = 20$ ; high:  $\geq 8$  at 20 weeks,  $n = 17$ ). Normal tail skin from FVB/BX, SPRET/Ei, FVB/N and F<sub>1</sub> mice (four replicates per group) was snap-frozen when mice were killed. Messenger RNA expression profiles were generated with the Affymetrix M430 2.0 platform. Mice were genotyped using the ABI platform at 223 single nucleotide polymorphisms. Correlation analysis (Wilcoxon rank-sum) was performed with 24,357 transcripts expressed above background levels in FVB/BX samples. Eighty-three additional FVB/BX mice were generated and CBC was analysed using a blood cell counter (HEMAVET, CDC Technologies). Linkage testing was performed by linear regression. An FDR method that accounts for signal dependence between loci on the same chromosome<sup>26</sup> identified 5% and 10% FDR  $P$ -value cutoffs for linkage. CBC QTL



**Figure 4 | The inflammation/barrier function gene expression and linkage network.** Drawn as in Fig. 2. These genes are associated with susceptibility to papillomas, and many share a common significant eQTL at chromosome 15, 101 Mb. See also Supplementary Table 9.



were calculated and plotted using R/QTL<sup>27</sup>, using 1,000 permutations to calculate the 5% genome-wide error rate *P*-value cutoff. Gene ontology analysis was performed with BiNGO<sup>28</sup>. The relevance network used an  $r^2$  cutoff of 0.64 (<0.01% alpha level, 1,000 permutations). Cliques with at least five members were identified using a modification of the Bron–Kerbosch algorithm. Network figures were generated using Cytoscape version 2.6.1 (www.cytoscape.org). Differentially expressed genes were identified using the Significance Analysis of Microarrays<sup>29</sup> with a maximum FDR of 10%.

Received 7 July; accepted 8 December 2008.

Published online 11 January 2009.

- Hubner, N. *et al.* Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. *Nature Genet.* **37**, 243–253 (2005).
- Emilsson, V. *et al.* Genetics of gene expression and its effect on disease. *Nature* **452**, 423–428 (2008).
- Chen, Y. *et al.* Variations in DNA elucidate molecular networks that cause disease. *Nature* **452**, 429–435 (2008).
- Nagase, H. *et al.* Distinct genetic loci control development of benign and malignant skin tumours in mice. *Nature Genet.* **10**, 424–429 (1995).
- Ewart-Toland, A. *et al.* Identification of *Stk6/STK15* as a candidate low-penetrance tumor-susceptibility gene in mouse and human. *Nature Genet.* **34**, 403–412 (2003).
- Manenti, G. *et al.* Genetic mapping of a pulmonary adenoma resistance (*Par1*) in mouse. *Nature Genet.* **12**, 455–457 (1996).
- Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature* **449**, 1003–1007 (2007).
- Jaks, V. *et al.* *Lgr5* marks cycling, yet long-lived, hair follicle stem cells. *Nature Genet.* **40**, 1291–1299 (2008).
- Blumberg, H. *et al.* Opposing activities of two novel members of the IL-1 ligand family regulate skin inflammation. *J. Exp. Med.* **204**, 2603–2614 (2007).
- Murphy, J. E., Morales, R. E., Scott, J. & Kupper, T. S. IL-1 alpha, innate immunity, and skin carcinogenesis: the effect of constitutive expression of IL-1 alpha in epidermis on chemical carcinogenesis. *J. Immunol.* **170**, 5697–5703 (2003).
- Lipworth, B. J. Phosphodiesterase-4 inhibitors for asthma and chronic obstructive pulmonary disease. *Lancet* **365**, 167–175 (2005).
- Mahieu, T. *et al.* The wild-derived inbred mouse strain SPRET/Ei is resistant to LPS and defective in IFN- $\beta$  production. *Proc. Natl Acad. Sci. USA* **103**, 2292–2297 (2006).
- Coussens, L. M. & Werb, Z. Inflammation and cancer. *Nature* **420**, 860–867 (2002).
- Lin, W. W. & Karin, M. A cytokine-mediated link between innate immunity, inflammation, and cancer. *J. Clin. Invest.* **117**, 1175–1183 (2007).
- Viaje, A., Slaga, T. J., Wigler, M. & Weinstein, I. B. Effects of antiinflammatory agents on mouse skin tumor promotion, epidermal DNA synthesis, phorbol ester-induced cellular proliferation, and production of plasminogen activator. *Cancer Res.* **37**, 1530–1536 (1977).
- Hoffjan, S. & Stemmler, S. On the role of the epidermal differentiation complex in ichthyosis vulgaris, atopic dermatitis and psoriasis. *Br. J. Dermatol.* **157**, 441–449 (2007).
- Koch, P. J. *et al.* Lessons from loricrin-deficient mice: compensatory mechanisms maintaining skin barrier function in the absence of a major cornified envelope protein. *J. Cell Biol.* **151**, 389–400 (2000).
- Weidinger, S. *et al.* Loss-of-function variations within the filaggrin gene predispose for atopic dermatitis with allergic sensitizations. *J. Allergy Clin. Immunol.* **118**, 214–219 (2006).
- Hollox, E. J. *et al.* Psoriasis is associated with increased  $\beta$ -defensin genomic copy number. *Nature Genet.* **40**, 23–25 (2008).
- Gebhardt, C., Nemeth, J., Angel, P. & Hess, J. S100A8 and S100A9 in inflammation and cancer. *Biochem. Pharmacol.* **72**, 1622–1631 (2006).
- Giovannucci, E. Vitamin D status and cancer incidence and mortality. *Adv. Exp. Med. Biol.* **624**, 31–42 (2008).
- Bikle, D. D. *et al.* The vitamin D response element of the involucrin gene mediates its regulation by 1,25-dihydroxyvitamin D<sub>3</sub>. *J. Invest. Dermatol.* **119**, 1109–1113 (2002).
- Schauber, J. *et al.* Injury enhances TLR2 function and antimicrobial peptide expression through a vitamin D-dependent mechanism. *J. Clin. Invest.* **117**, 803–811 (2007).
- Cianferotti, L., Cox, M., Skorija, K. & Demay, M. B. Vitamin D receptor is essential for normal keratinocyte stem cell function. *Proc. Natl Acad. Sci. USA* **104**, 9428–9433 (2007).
- Zinser, G. M., Sundberg, J. P. & Welsh, J. Vitamin D<sub>3</sub> receptor ablation sensitizes skin to chemically induced tumorigenesis. *Carcinogenesis* **23**, 2103–2109 (2002).
- Chen, L. & Storey, J. D. Relaxed significance criteria for linkage analysis. *Genetics* **173**, 2371–2381 (2006).
- Broman, K. W., Wu, H., Sen, S. & Churchill, G. A. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**, 889–890 (2003).
- Maere, S., Heymans, K. & Kuiper, M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **21**, 3448–3449 (2005).
- Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA* **98**, 5116–5121 (2001).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We are grateful to R. Del Rosario for technical assistance with mouse breeding. We thank G. Hirst, R. Akhurst and H. Quigley for their comments. This work was supported by the National Cancer Institute. A.B. acknowledges support from the Barbara Bass Bakar Chair of Cancer Genetics. M.D.T. was supported in part by a Sandler Foundation postdoctoral research fellowship. J.P.-L. is an investigator of the 'Programa Ramón y Cajal' from the Spanish 'Ministerio de Educación y Ciencia' partially supported by the European Community; his research is partially funded by the 'Fondo de Investigaciones Sanitarias' and Junta de Castilla y León.

**Author Contributions** The study was conceived and supervised by A.B. The software was written and the bioinformatics analysis was carried out by D.A.Q. H.N. carried out the primary tumour induction experiments, and M.D.T. isolated the DNA and RNA and carried out the gene expression microarray analysis. D.G.G. provided the genotyping data, J.-H.M. provided statistical support, and F.G.P. performed Tagman validation and DNA sequencing. J.P.-L. carried out the separate backcross for analysis of blood parameters, and measured all blood phenotypes. The paper was written by D.A.Q. and A.B., with important contributions from the other authors.

**Author Information** Microarray results have been deposited in GEO under accession number GSE12248. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to A.B. (abalmain@cc.ucsf.edu).

# AIM2 activates the inflammasome and cell death in response to cytoplasmic DNA

Teresa Fernandes-Alnemri<sup>1\*</sup>, Je-Wook Yu<sup>1\*</sup>, Pinaki Datta<sup>1</sup>, Jianghong Wu<sup>1</sup> & Emad S. Alnemri<sup>1</sup>

Host- and pathogen-associated cytoplasmic double-stranded DNA triggers the activation of a NALP3 (also known as cryopyrin and NLRP3)-independent inflammasome<sup>1</sup>, which activates caspase-1 leading to maturation of pro-interleukin-1 $\beta$  and inflammation. The nature of the cytoplasmic-DNA-sensing inflammasome is currently unknown. Here we show that AIM2 (absent in melanoma 2), an interferon-inducible HIN-200 family member that contains an amino-terminal pyrin domain and a carboxy-terminal oligonucleotide/oligosaccharide-binding domain<sup>2,3</sup>, senses cytoplasmic DNA by means of its oligonucleotide/oligosaccharide-binding domain and interacts with ASC (apoptosis-associated speck-like protein containing a CARD) through its pyrin domain to activate caspase-1. The interaction of AIM2 with ASC also leads to the formation of the ASC pyroptosome<sup>4</sup>, which induces pyroptotic cell death in cells containing caspase-1. Knockdown of AIM2 by short interfering RNA reduced inflammasome/pyroptosome activation by cytoplasmic DNA in human and mouse macrophages, whereas stable expression of AIM2 in the non-responsive human embryonic kidney 293T cell line conferred responsiveness to cytoplasmic DNA. Our results show that cytoplasmic DNA triggers formation of the AIM2 inflammasome by inducing AIM2 oligomerization. This study identifies AIM2 as an important inflammasome component that senses potentially dangerous cytoplasmic DNA, leading to activation of the ASC pyroptosome and caspase-1.

A key innate immune response against infection with microbial or viral pathogens and tissue damage is the rapid activation of multi-protein complexes called inflammasomes<sup>5</sup>. The inflammasomes activate caspase-1, a cysteine protease that processes the inactive pro-interleukin-1 $\beta$  (pro-IL1 $\beta$ ) and pro-IL18 to their active pro-inflammatory cytokines IL1 $\beta$  and IL18, respectively. The most studied among these is the NALP3 inflammasome<sup>5</sup>, which is activated by diverse stimuli perhaps by a lysosomal destabilization mechanism<sup>6,7</sup>.

A recent study showed that DNA from different sources activates an ASC-dependent, but a NALP3-independent, inflammasome<sup>1</sup>. To identify the DNA-sensing inflammasome, we searched the NCBI database for proteins with pyrin and oligonucleotide-binding domains. We identified four human proteins (IFI16, AIM2, IFIX and MNDA), which belong to the interferon-inducible HIN-200 family<sup>2,3</sup>, that meet these two criteria. Investigation of the ability of these proteins to activate caspase-1 when ectopically expressed in the stable 293T-caspase-1-ASC cell line<sup>8</sup> (293T cell line containing caspase-1 and ASC) showed that AIM2 is the only member of the HIN-200 family capable of activating caspase-1 (Fig. 1a and Supplementary Fig. 1). The activation of caspase-1 by AIM2 was dependent on an intact pyrin domain (PYD) because deletion of the PYD of AIM2 completely abrogated caspase-1 activation (Fig. 1b). This was also dependent on ASC, because no AIM2-induced caspase-1 activation was observed

in the 293T-caspase-1 cells<sup>8</sup>, which lack ASC (Fig. 1a and Supplementary Fig. 1a). Additionally, expression of AIM2 induced secretion of activated caspase-1 and IL1 $\beta$  from stable 293T-caspase-1-ASC-pro-IL1 $\beta$  cells, which express ASC, but not from 293T-caspase-1-pro-IL1 $\beta$  cells, which lack ASC (Fig. 1c). Together, these results indicate that AIM2 can activate caspase-1 to produce the active IL1 $\beta$  cytokine in an ASC-dependent manner, perhaps by engaging ASC and inducing its oligomerization. Indeed, expression of AIM2 in 293T-ASC-EGFP-N1 cells<sup>8</sup> (293T cell line containing enhanced green fluorescent protein (EGFP)-tagged ASC) resulted in the formation of the oligomeric ASC pyroptosome that we showed recently to participate in caspase-1 activation during pyroptosis<sup>4</sup> (Fig. 1d and Supplementary Fig. 2).

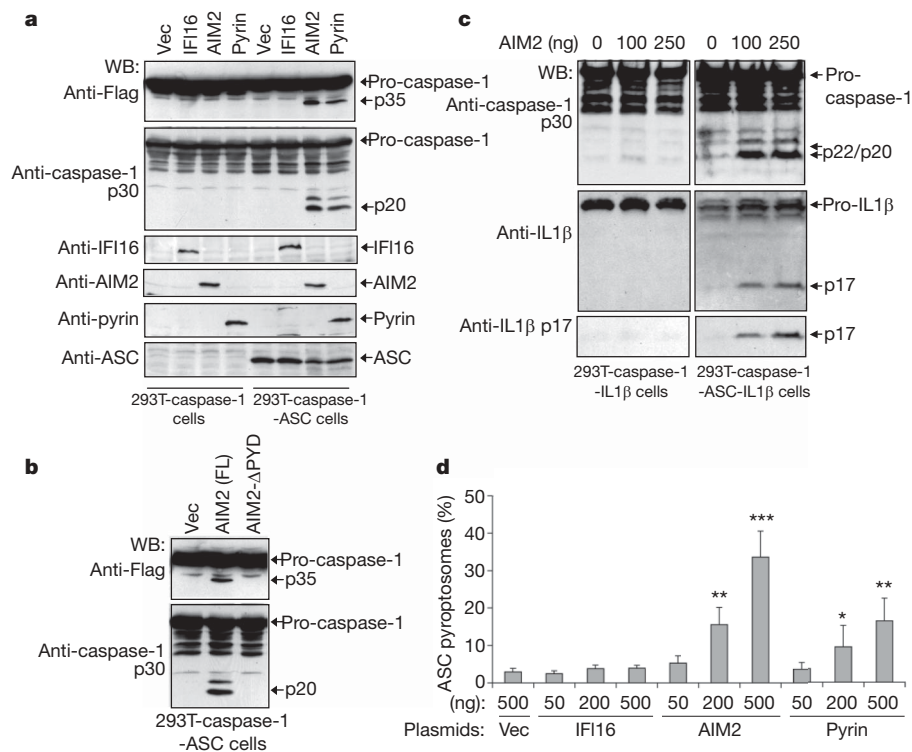
To gain some insight into how AIM2 induces ASC-dependent caspase-1 activation, we tested whether AIM2 interacts directly with ASC to activate caspase-1. *In vitro* pull-down experiments revealed that full-length AIM2, but not a truncated AIM2 lacking the PYD (AIM2- $\Delta$ PYD), can interact with ASC and its isolated PYD (Fig. 2a, b). Furthermore, immunoprecipitation of endogenous ASC from interferon- $\gamma$ -induced human THP-1 cell lysates with an ASC-specific antibody resulted in precipitation of endogenous AIM2 (Fig. 2c). No detectable interaction between ASC and endogenous IFI16, MNDA or IFIX was observed (Fig. 2c, fourth to sixth panels from the top), further supporting our earlier conclusion that only AIM2, among members of the HIN-200 family, can specifically interact with ASC to activate caspase-1. Together, these results indicate that AIM2 associates directly with ASC by means of PYD–PYD interactions. This interaction is reminiscent of the interaction of NALP3 or pyrin with the PYD of ASC, which induces ASC oligomerization leading to activation of caspase-1 in response to various stimuli<sup>8–10</sup>.

Structural analysis of the C-terminal HIN-200 domain of AIM2 revealed that this domain contains two adjacent oligonucleotide/oligosaccharide-binding folds<sup>11</sup> that could potentially interact with DNA. Electrophoretic mobility shift assays with purified HIN-200 domain of AIM2 (AIM2- $\Delta$ PYD; Supplementary Fig. 4a, left panel) and the synthetic DNA poly(dA:dT)•poly(dA:dT) (referred to throughout as poly(dA:dT)) showed formation of increasing amounts of slow-migrating AIM2–DNA complexes with increasing amounts of AIM2- $\Delta$ PYD (Supplementary Fig. 4a, right panel), indicating that this domain can indeed bind to DNA. As expected, a purified full-length AIM2, but not the isolated PYD, was also capable of binding to DNA and inducing a gel-shift of poly(dA:dT) (Supplementary Fig. 4b).

Given that AIM2 can interact with ASC and DNA through its PYD and HIN-200 domains, respectively, we next used AIM2-targeting short interfering RNA (siRNA) to demonstrate that AIM2 has an important role in caspase-1 activation by cytoplasmic DNA. Knocking down human AIM2 in THP-1 cells clearly reduced

<sup>1</sup>Department of Biochemistry and Molecular Biology, Center for Apoptosis Research, Kimmel Cancer Institute, Thomas Jefferson University, Philadelphia, Pennsylvania 19107, USA.

\*These authors contributed equally to this work.



**Figure 1 | Activation of caspase-1 by AIM2.** **a, b**, The indicated cells ( $1 \times 10^6$  cells per 35-mm well) were transfected with 0.5  $\mu$ g of an empty vector (Vec) or constructs encoding human AIM2, IFI16 or pyrin (**a**), or with Vec, full-length (FL) AIM2 or AIM2- $\Delta$ PYD plasmids (**b**), as indicated for 24 h. Lysates were analysed by western blotting (WB) with anti-Flag antibody (first panel from the top) or anti-caspase-1 p30 polyclonal antibody (second panel). The blot in **a** was re-probed with IFI16, AIM2,

poly(dA:dT)-induced caspase-1 activation and IL1 $\beta$  processing (Fig. 2d, fourth lane, and Supplementary Fig. 5a, b). Knocking down AIM2 had little effect on caspase-1 and IL1 $\beta$  activation by monosodium urate (MSU), which activates caspase-1 through the NALP3 inflammasome (Fig. 2d, fifth and sixth lanes). Consistent with these results, knocking down mouse AIM2 in immortalized wild-type and *NALP3*<sup>-/-</sup> mouse bone marrow macrophages also significantly reduced poly(dA:dT)-induced caspase-1 activation (Fig. 2e and Supplementary Fig. 6a–c). The *Aim2*-targeting siRNAs had little effect on caspase-1 activation by lipopolysaccharide (LPS) plus ATP in mouse wild-type macrophages (Supplementary Fig. 6d and Supplementary Note 1). Together, these results indicate that AIM2 in human and mouse macrophages senses cytoplasmic DNA through its HIN-200 domain and is critical for caspase-1 activation by cytoplasmic DNA.

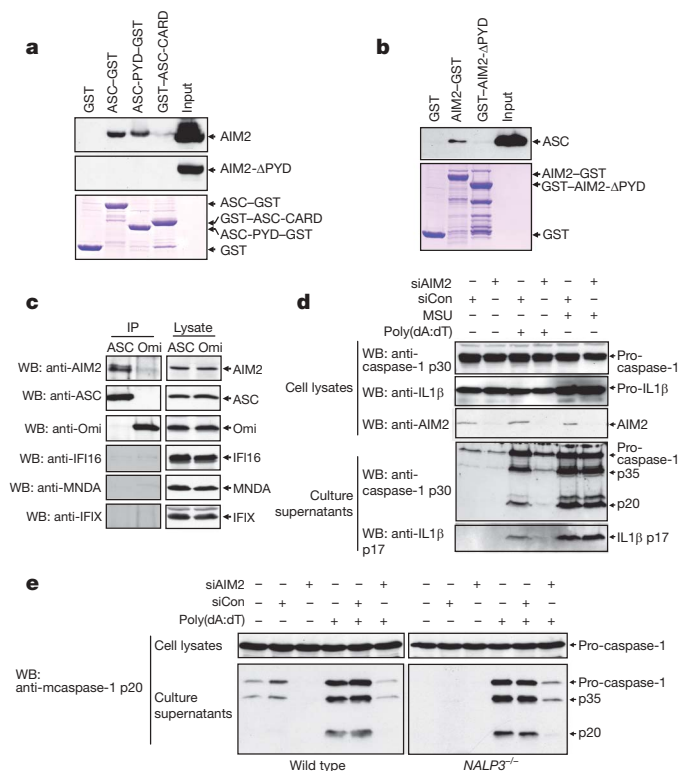
A pyroptosome formation assay in THP-1-ASC-EGFP-N1 cells<sup>4</sup> revealed that cytoplasmic DNA induces robust ASC oligomerization (Fig. 3a). To provide direct evidence that AIM2 is required for DNA-induced ASC oligomerization, we used 293T cells, which do not express endogenous AIM2 (Supplementary Fig. 7a), to generate a stable 293T cell line that co-expresses AIM2 and ASC-EGFP proteins (designated 293T-AIM2-ASC-EGFP-N1). Transfection with poly(dA:dT) induced ASC pyroptosome formation in 293T-AIM2-ASC-EGFP-N1 cells within 2 h after transfection (Fig. 3b, lower right panel), but not in control 293T-ASC-EGFP-N1 cells, which do not express AIM2 (Fig. 3b, lower left panel). Quantification of the ASC pyroptosomes revealed a dose-dependent effect of poly(dA:dT) in these cells (Fig. 3c). Consistent with these results, DNA from different sources including poly(dA:dT) (Supplementary Note 2) was able to induce caspase-1 activation in a stable 293T-caspase-1-ASC-AIM2 cell line that co-expresses pro-caspase-1, ASC and AIM2, but not in the parental 293T-caspase-1-ASC cell line, which lacks AIM2

pyrin or ASC antibodies (third to sixth panels, respectively). **c**, Immunoblots for caspase-1 (top panel), IL1 $\beta$  (middle panel) and IL1 $\beta$  p17 (bottom panel) in culture supernatants of the indicated cell lines transfected with the indicated amounts of AIM2 expression plasmid. **d**, Percentages of ASC pyroptosomes in 293T-ASC-EGFP-N1 cells after transfection with the indicated plasmids. Values represent mean and s.d. ( $n = 4$ ); \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.005$ .

(Fig. 3d). The expression level of AIM2 in the 293T-caspase-1-ASC-AIM2 cells is comparable to the expression of endogenous AIM2 in interferon- $\gamma$ -induced THP-1 cells (Fig. 3e). Transfection of plasmid DNA into a stable 293T-caspase-1-ASC-NALP3 cell line that co-expresses pro-caspase-1, ASC and NALP3 did not induce caspase-1 activation, indicating that, unlike AIM2, NALP3 does not participate directly in sensing cytoplasmic DNA (Fig. 3f). These results underscore the essential role of AIM2 in cytoplasmic DNA sensing and DNA-induced ASC oligomerization and caspase-1 activation, and indicate that binding of cytoplasmic DNA to AIM2 generates the molecular signal necessary for this process to occur.

To understand how cytoplasmic DNA induces AIM2 activation, we tested the ability of purified recombinant human AIM2 to induce activation of a chimaeric PYD-caspase-1 protein, which contains an N-terminal PYD derived from ASC instead of the CARD, *in vitro* (see Supplementary Note 3). Incubation of purified AIM2 or ASC pyroptosome with PYD-caspase-1 resulted in its activation (Fig. 4a, second and fourth lanes, respectively). No activation of PYD-caspase-1 was seen when it was incubated with AIM2- $\Delta$ PYD, which lacks the PYD (third lane), indicating that the interactions of the PYD of AIM2 with the PYD of PYD-caspase-1 is required for activation. AIM2 was also able to activate a chimaeric wild-type PYD-caspase-9, but not an active site mutant PYD-caspase-9(C287A) (Fig. 4b). Because initiator pro-caspases are activated by dimerization, these results indicate that AIM2 activates the pro-caspase-1 and pro-caspase-9 chimaeras by inducing the dimerization of their inactive monomeric forms. Unexpectedly, incubation of AIM2 with poly(dA:dT) did not increase the ability of AIM2 to activate PYD-caspase-9 or PYD-caspase-1 (Fig. 4c and data not shown). This indicates that the purified AIM2 might have already been activated by bacterial DNA during expression in bacteria or self-oligomerized during purification. Indeed, chemical cross-linking experiments in



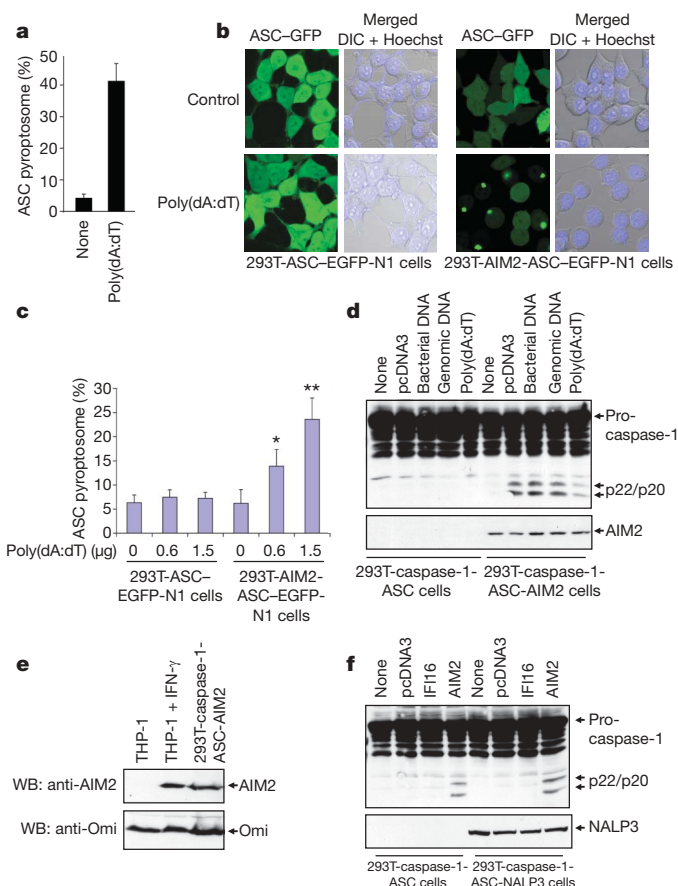


**Figure 2 | AIM2 interacts with ASC to activate caspase-1.**

**a, b**, Immunoblots showing *in vitro* association of ASC–GST (glutathione S-transferase) and ASC–PYD–GST with full-length AIM2 (top panel), but not with truncated AIM2– $\Delta$ PYD (middle panel) (**a**), and *in vitro* association of AIM2–GST with full-length ASC (**b**). The bottom panels show Coomassie-stained gel of the GST-tagged proteins used in the pull-down assay (see Supplementary Fig. 3). GST–ASC–CARD refers to the isolated CARD of ASC fused to GST. **c**, Association of endogenous ASC with endogenous AIM2 in THP-1 cell lysate detected by immunoprecipitation (IP, first lanes, left panels) with an ASC-specific monoclonal antibody as described in the Methods. IP with an Omi-specific monoclonal antibody was used as a negative control (second lanes, left panels). **d, e**, Immunoblots of lysates and culture supernatants of THP-1 cells (**d**) or mouse wild-type and *NALP3*<sup>−/−</sup> bone marrow macrophages (**e**) transfected with AIM2-targeting siRNA (siAIM2) or control non-specific siRNA (siCon) followed by transfection with poly(dA:dT) or treatment with MSU as indicated. mcaspase-1, mouse caspase-1. The release of the active caspase-1 p20 subunit and some pro-caspase-1 in the culture supernatants is a result of cell death and lysis (see Fig. 5 and Supplementary Movie 1).

the absence of DNA showed that the purified AIM2 preparation contains dimeric and oligomeric forms of AIM2 (Fig. 4d, second lane). Interestingly, cross-linking in the presence of DNA increased the oligomeric forms of AIM2 (Fig. 4d, third and fourth lanes) in a dose-dependent manner, as evidenced by the formation of very large AIM2 oligomers that failed to enter the gel. This result indicates that binding of double-stranded DNA to AIM2 induces its oligomerization, which explains the ability of cytoplasmic DNA to activate the AIM2 inflammasome in macrophages and in AIM2-reconstituted 293T cells.

To provide direct evidence that cytoplasmic DNA can indeed induce AIM2 oligomerization, we transfected 293T–AIM2–EGFP–N1 cells or *NALP3*<sup>−/−</sup>–AIM2–EGFP–N1 mouse macrophages cells that stably express human or mouse EGFP-tagged AIM2, respectively, with a Cy3-labelled plasmid DNA. Confocal live cell imaging revealed that DNA induced notable clustering of the cytoplasmic AIM2–EGFP around the transfected DNA in both 293T cells and *NALP3*<sup>−/−</sup> macrophages (Fig. 5b, d and Supplementary Fig. 9). DNA did not induce clustering of free EGFP in the control 293T–EGFP cells (Fig. 5a), indicating that AIM2 is responsible for the clustering of AIM2–EGFP in the 293T and *NALP3*<sup>−/−</sup> macrophages.

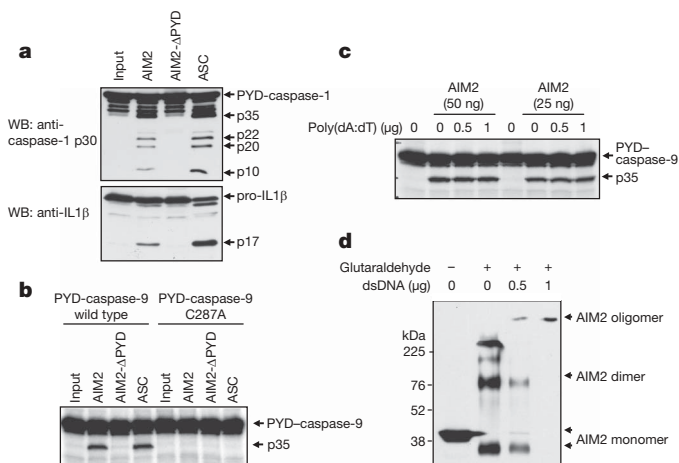


**Figure 3 | Cell-based reconstitution of the AIM2 inflammasome.**

**a**, Percentages of ASC pyroptosomes in untransfected (none) or poly(dA:dT)-transfected THP-1 cells. Values represent mean and s.d. ( $n = 3$ ). **b**, Fluorescence confocal images showing formation of the speck-like ASC pyroptosomes only in 293T–AIM2–ASC–EGFP–N1 cells (right panels), but not in 293T–ASC–EGFP–N1 cells (left panels), 2 h after transfection with vehicle (control, top panels) or poly(dA:dT) (bottom panels). DIC, differential interference contrast. **c**, Percentages of ASC pyroptosomes in the indicated cell lines after transfection with poly(dA:dT) for 2 h. Values represent mean and s.d. ( $n = 7$ ); \* $P < 0.005$ , \*\* $P < 0.001$ . **d**, Immunoblot for caspase-1 in lysates from 293T–caspase-1–ASC (lanes 1–5) or 293T–caspase-1–ASC–AIM2 (lanes 6–10) cells after transfection with the indicated types of DNA (0.5  $\mu$ g per  $1 \times 10^6$  cells) for 24 h. **e**, Immunoblot for AIM2 (top panel) or Omi (bottom panel) in lysates from uninduced (lane 1) or interferon- $\gamma$ -induced (IFN- $\gamma$ , lane 2) THP-1 cells, or 293T–caspase-1–ASC–AIM2 cells (lane 3). **f**, Immunoblot for caspase-1 in lysates from 293T–caspase-1–ASC (lanes 1–4) or 293T–caspase-1–ASC–*NALP3* (lanes 5–8) cells after transfection with the indicated plasmid DNA (0.5  $\mu$ g per  $1 \times 10^6$  cells) for 24 h.

Interestingly, DNA-induced clustering of AIM2–EGFP caused pyroptotic cell death with characteristic plasma membrane swelling in the *NALP3*<sup>−/−</sup>–AIM2–EGFP–N1 cells (Fig. 5d, Supplementary Fig. 10 and Supplementary Movie 1), but not in the 293T–AIM2–EGFP–N1 cells (Fig. 5b), because these cells lack ASC and caspase-1. Pretreatment of the *NALP3*<sup>−/−</sup>–AIM2–EGFP–N1 cells with the caspase inhibitor zVAD-FMK (Z-Val-Ala-Asp-Fluoromethyl ketone) did not affect DNA-induced clustering of AIM2–EGFP in these cells, but prevented cell death (Supplementary Fig. 9b). Taken together, these results provide direct evidence that cytoplasmic DNA binds to AIM2 and induces its oligomerization in live cells. This is the first demonstration of an inflammasome bound to its ligand in live cells.

The putative tumour-suppressive activity of AIM2 (ref. 12) might be attributed to its ability to induce pyroptosis in cells that co-express ASC and pro-caspase-1. Indeed, transient expression of AIM2 in 293T–ASC–caspase-1 cells resulted in pyroptotic cell death as measured by the release of lactate dehydrogenase (LDH) in the culture medium



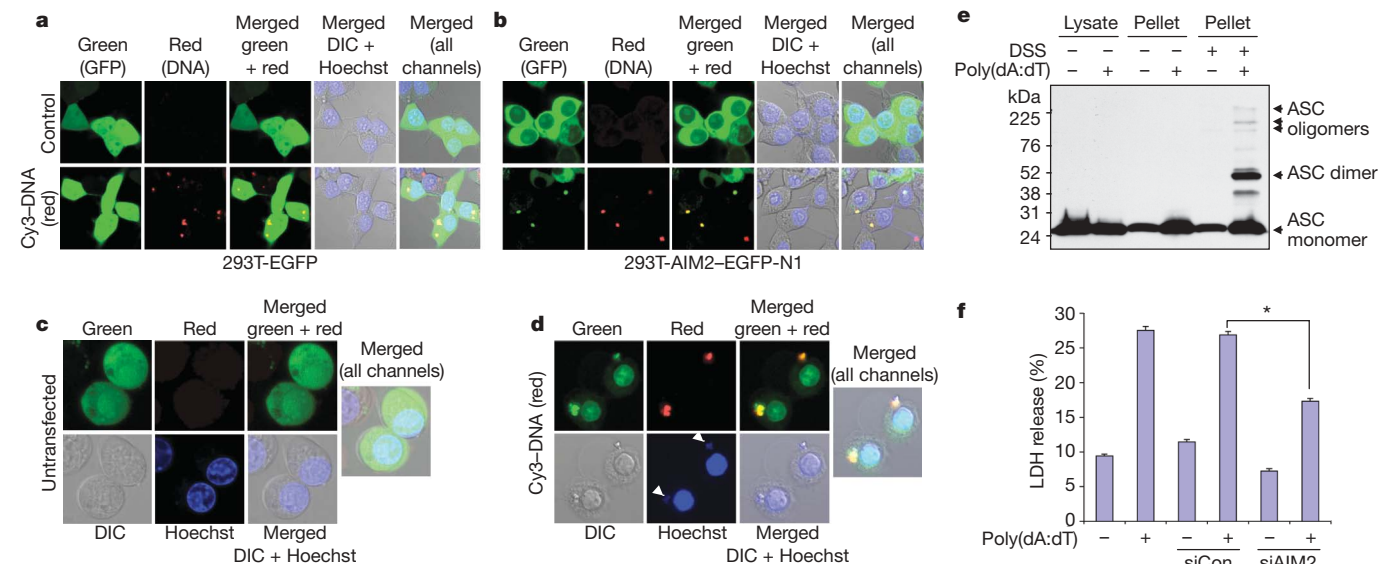
**Figure 4 | In vitro reconstitution of the AIM2 inflammasome.**

**a**, Immunoblot for caspase-1 (top panel) or IL1β (bottom panel) after *in vitro* incubation of PYD-caspase-1 and pro-IL1β with purified GST-AIM2 (lane 2), GST-AIM2-ΔPYD (lane 3) or ASC pyroptosome (lane 4). **b, c**, Autoradiograms of <sup>35</sup>S-labelled wild-type or C287A chimaeric PYD-caspase-9 after *in vitro* incubation with purified GST-AIM2 (lanes 2 and 6), GST-AIM2-ΔPYD (lanes 3 and 7) or ASC pyroptosome (lanes 4 and 8) (**b**) or with the indicated amounts of purified GST-AIM2 in the presence or absence of the indicated amounts of poly(dA:dT) (**c**). **d**, Immunoblot for AIM2 after cross-linking of purified AIM2 (2 μg) with glutaraldehyde in the presence of the indicated amounts of 64-mer double-stranded (ds)DNA. A longer exposure of this gel is shown in Supplementary Fig. 8.

(Supplementary Fig. 11a). No AIM2-induced cell death was observed in cells that lack caspase-1 or ASC (such as 293T-caspase-1, 293T-ASC-EGFP-N1 or 293T-AIM2-EGFP-N1 cells; Fig. 5b, Supplementary Fig. 11a and data not shown). Consistent with this, transfection of poly(dA:dT) or plasmid DNA into the *NALP3*<sup>-/-</sup> bone marrow

macrophages induced ASC pyroptosome formation (Fig. 5e), cell death with characteristic features of pyroptosis (Supplementary Fig. 11b, lower panels, and Supplementary Movie 1) and release of LDH into the culture medium (Fig. 5f). Knocking down AIM2 by *Aim2*-specific siRNA significantly reduced cell death and LDH release in these cells (Fig. 5f). These data indicate that AIM2 is a bona fide activator of inflammation and caspase-1-dependent cell death (pyroptosis) in cells that express both ASC and caspase-1.

Our work clearly identifies AIM2 as a major activator of caspase-1 in response to cytoplasmic DNA. As demonstrated *in vitro* and in live cells, DNA binds directly to AIM2 and induces its oligomerization. We suggest that the oligomeric AIM2-DNA complex serves as a molecular platform, much like the oligomeric NALP3 platform, to recruit ASC and facilitate the self-association of its PYD to form ASC dimers, which subsequently oligomerize with other ASC dimers to form a large structure we called recently the 'ASC pyroptosome'<sup>4</sup>. This is supported by our observation that engagement of AIM2 by cytoplasmic DNA leads to the formation of ASC pyroptosomes in THP-1 and *NALP3*<sup>-/-</sup> macrophages and in 293T-AIM2-ASC-EGFP-N1 cells (Figs 3 and 5). Cytoplasmic DNA, whether derived from infection with viral or microbial pathogens, tissue damage or delivered by cationic liposomes, is a strong activator of type I interferon response. AIM2 is an interferon-inducible protein and we have observed upregulation of AIM2 in THP-1 macrophages after treatment with interferon-γ, LPS or transfection with the synthetic DNA poly(dA:dT). Our data indicate that the induction of AIM2 serves as a backup innate immune response to the NALP3 inflammasome, capable of sensing cytoplasmic DNA and of triggering a strong pro-inflammatory response. The perceived redundancy in the response to DNA by the NALP3 and AIM2 inflammasomes could be explained by differences in the mechanisms by which these two inflammasomes are activated. Recent studies indicate that the NALP3 inflammasome is activated by many stimuli that cause phagosomal/lysosomal destabilization<sup>6,7,13-15</sup>. When DNA viruses (for example, adenovirus) infect macrophages, the



**Figure 5 | Cytoplasmic DNA-induced AIM2 oligomerization and pyroptosis.**

**a, b**, Confocal live cell images of 293T-EGFP (a) or 293T-AIM2-EGFP-N1 (b) cells after transfection with vehicle (control, top panels) or Cy3-labelled DNA (red, bottom panels). **c, d**, Confocal live cell images of *NALP3*<sup>-/-</sup>-AIM2-EGFP-N1 bone marrow macrophages after transfection with vehicle (**c**) or Cy3-labelled DNA (**d**). Notice the oligomerization of AIM2-EGFP by the red DNA in the 293T-AIM2-EGFP-N1 cells (**b**, bottom panels) and *NALP3*<sup>-/-</sup>-AIM2-EGFP-N1 macrophages (**d**). The green (AIM2-EGFP), red (Cy3-labelled DNA), grey (DIC) and blue (Hoechst stain, nucleus) channels, and the merged channels, are indicated. The two white arrows in **d** indicate staining of the cytoplasmic DNA with the blue Hoechst stain, which specifically stains DNA. Notice the pyroptotic cell

death features induced by the Cy3-labelled DNA in *NALP3*<sup>-/-</sup>-AIM2-EGFP-N1 macrophages (**d**), but not in the 293T-AIM2-EGFP-N1 (**b**). **e**, Immunoblot for ASC showing the presence of the oligomeric ASC pyroptosomes only in the pellets of poly(dA:dT)-transfected (lane 6), but not in the pellets of untransfected (lane 5), *NALP3*<sup>-/-</sup> macrophages. A shorter exposure and a more detailed explanation of this gel is shown in Supplementary Fig. 12. DSS, disuccinimidyl suberate. **f**, Percentages of LDH release into the culture medium of *NALP3*<sup>-/-</sup> macrophages transfected with vehicle (columns 1 and 2), non-specific control siRNA (siCon, columns 3 and 4) or mouse *Aim2*-specific siRNA (siAIM2, columns 5 and 6) followed by transfection with or without poly(dA:dT) (1 μg per 1 × 10<sup>6</sup> cells) as indicated. Values represent mean and s.d. (n = 3); \*P < 0.001.

DNA-containing viral particles accumulate in the phagosomes<sup>16</sup>, which could lead to destabilization of the phagosome thereby resulting in activation of NALP3 as has recently been shown<sup>1</sup>. In contrast, when DNA escapes the phagosomes it could be sensed directly by AIM2, which is induced by the initial type I interferon response to viral entry, by means of direct interaction with its oligonucleotide/oligosaccharide-binding-fold-containing domain, thereby leading to AIM2 oligomerization and activation. It is therefore probable that the NALP3 and AIM2 inflammasomes represent two important lines of defence to ensure a full pro-inflammatory response to potentially dangerous cytoplasmic DNA. Additional discussion is provided in the Supplementary Discussion.

## METHODS SUMMARY

**Assay of caspase-1 activation.** All cells were seeded in 35-mm 6-well plates. Stimulation with DNA (poly(dA:dT), plasmid, genomic or bacterial) were performed in serum-free (SF) OPTI-MEM I by cationic liposome transfection using Lipofectamine 2000 (3.5 µl ml<sup>-1</sup>), as per the manufacturer's protocol (Invitrogen). After stimulation, cell culture supernatants and cell pellets were analysed by immunoblotting with anti-caspase-1- and anti-IL1β-specific antibodies.

**siRNA knockdown.** THP-1 cells, or immortalized wild-type or *NALP3*<sup>-/-</sup> macrophages, were attached to 35-mm 6-well plates. Cells were then transfected with siRNA oligonucleotides (50 nM) for 24 h using Lipofectamine 2000 in 2 ml of SF OPTI-MEM I. The next day, cells were washed with SF OPTI-MEM I and then transfected with poly(dA:dT) (1 µg ml<sup>-1</sup>) for 6 h with Lipofectamine 2000 in 2 ml of SF OPTI-MEM I. The culture supernatants and cells were separated and then processed for immunoblot analysis with anti-caspase-1- and anti-IL1β-specific antibodies.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 11 November; accepted 5 December 2008.

Published online 21 January 2009.

1. Muruve, D. A. *et al.* The inflammasome recognizes cytosolic microbial and host DNA and triggers an innate immune response. *Nature* **452**, 103–107 (2008).

2. Choubey, D. & Panchanathan, R. Interferon-inducible Ifi200-family genes in systemic lupus erythematosus. *Immunol. Lett.* **119**, 32–41 (2008).

3. Ludlow, L. E., Johnstone, R. W. & Clarke, C. J. The HIN-200 family: more than interferon-inducible genes? *Exp. Cell Res.* **308**, 1–17 (2005).

4. Fernandes-Alnemri, T. *et al.* The pyroptosome: a supramolecular assembly of ASC dimers mediating inflammatory cell death via caspase-1 activation. *Cell Death Differ.* **14**, 1590–1604 (2007).

5. Petrilli, V., Dostert, C., Muruve, D. A. & Tschopp, J. The inflammasome: a danger sensing complex triggering innate immunity. *Curr. Opin. Immunol.* **19**, 615–622 (2007).
6. Halle, A. *et al.* The NALP3 inflammasome is involved in the innate immune response to amyloid-beta. *Nature Immunol.* **9**, 857–865 (2008).
7. Hornung, V. *et al.* Silica crystals and aluminum salts activate the NALP3 inflammasome through phagosomal destabilization. *Nature Immunol.* **9**, 847–856 (2008).
8. Yu, J. W. *et al.* Pyrin activates the ASC pyroptosome in response to engagement by autoinflammatory PSTPIP1 mutants. *Mol. Cell* **28**, 214–227 (2007).
9. Yu, J. W. *et al.* Cryopyrin and pyrin activate caspase-1, but not NF-κB, via ASC oligomerization. *Cell Death Differ.* **13**, 236–249 (2006).
10. Agostini, L. *et al.* NALP3 forms an IL-1β-processing inflammasome with increased activity in Muckle-Wells autoinflammatory disorder. *Immunity* **20**, 319–325 (2004).
11. Albrecht, M., Choubey, D. & Lengauer, T. The HIN domain of IFI-200 proteins consists of two OB folds. *Biochem. Biophys. Res. Commun.* **327**, 679–687 (2005).
12. DeYoung, K. L. *et al.* Cloning a novel member of the human interferon-inducible gene family associated with control of tumorigenicity in a model of human melanoma. *Oncogene* **15**, 453–457 (1997).
13. Dostert, C. *et al.* Innate immune activation through Nalp3 inflammasome sensing of asbestos and silica. *Science* **320**, 674–677 (2008).
14. Eisenbarth, S. C., Colegio, O. R., O'Connor, W., Sutterwala, F. S. & Flavell, R. A. Crucial role for the Nalp3 inflammasome in the immunostimulatory properties of aluminium adjuvants. *Nature* **453**, 1122–1126 (2008).
15. Franchi, L. & Nunez, G. The Nlrp3 inflammasome is critical for aluminium hydroxide-mediated IL-1β secretion but dispensable for adjuvant activity. *Eur. J. Immunol.* **38**, 2085–2089 (2008).
16. Meier, O., Gastaldelli, M., Boucke, K., Hemmi, S. & Greber, U. F. Early steps of clathrin-mediated endocytosis involved in phagosomal escape of Fcγ receptor-targeted adenovirus. *J. Virol.* **79**, 2604–2613 (2005).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was supported by NIH grants AG14357 and AR055398 to E.S.A. We thank M. McCormick and D. Wang for technical assistance, X. Jiao for help with the confocal microscopy, J. Sagara for the anti-human ASC antibody, J. Yuan for the anti-mouse caspase-1 antibody, R. Johnstone for the anti-human AIM2 antibody, M.-C. Hung for the anti-human IFI antibody, T. Ouchi for the *IFI16* complementary DNA and E. Latz for the immortalized mouse bone marrow macrophages.

**Author Contributions** T.F.-A. and J.-W.Y. performed most of the experiments. J.W. and P.D. performed additional experiments and provided technical assistance with equal contribution. E.S.A. conceived and designed the experiments, analysed and interpreted the data (with T.F.-A. and J.-W.Y.), and directed the whole project.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to E.S.A. (E\_Alnemri@lac.jci.tju.edu).



## METHODS

**Cell culture.** Human 293T and THP-1 cell lines were purchased from ATCC. 293T cells were maintained in DMEM/F12 (GIBCO) supplemented with 10% FBS and 100 U ml<sup>-1</sup> penicillin and streptomycin. THP-1 cells were grown in RPMI 1640 (GIBCO) supplemented with 10% FBS, 2 mM L-glutamine, 10 mM HEPES, 1 mM sodium pyruvate, 1.5 g l<sup>-1</sup> sodium bicarbonate, 0.05 mM 2-mercaptoethanol and 100 U ml<sup>-1</sup> penicillin and streptomycin. The *v-myc* and *v-raf* immortalized mouse wild-type, *ASC*<sup>-/-</sup> and *NALP3*<sup>-/-</sup> macrophages (gift from E. Latz) were grown in DMEM (GIBCO) supplemented with 10% FBS, 100 U ml<sup>-1</sup> penicillin and streptomycin. All cells were grown at 37 °C with 5% CO<sub>2</sub>. Before stimulation, THP-1 cells were differentiated for 3 h with 0.5–1 μM PMA, washed and allowed to attach to culture dishes for 24 h. All experiments for immunoblot analysis of cell culture supernatants used SF OPTI-MEM I medium (GIBCO). Transfection experiments with the synthetic poly(dA:dT) DNA or other types of DNA (plasmid, genomic or bacterial) were performed using Lipofectamine 2000 (3.5 μl ml<sup>-1</sup>) as per the manufacturer's protocol (Invitrogen).

**Antibodies and reagents.** Mouse polyclonal anti-human AIM2 antibody was purchased from Abnova. This antibody can detect overexpressed AIM2 but not endogenous AIM2. The mouse monoclonal anti-human AIM2 antibody (3B10) used to detect endogenous AIM2 in THP-1 cell lysates was obtained from R. Johnstone. The anti-human IFI16 monoclonal (sc-8023) and anti-human MND A polyclonal (sc-6051) antibodies were obtained from Santa Cruz Biotechnology. The anti-human IFIX polyclonal antibody was obtained from M.-C. Hung. The anti-IL1β monoclonal antibody (32D) was obtained from the NCI preclinical repository, biological resource branch. Rabbit polyclonal anti-human mature (17 kDa) IL1β (Asp 116) was obtained from Cell Signaling. Other antibodies used against human ASC (anti-human ASC monoclonal; from J. Sagara), human caspase-1 p30 (anti-human caspase-1, residues 100–404; our laboratory) and mouse caspase-1 p20 (anti-mouse caspase-1 p20; from J. Yuan) have been described previously<sup>4,8</sup>. Poly(dA:dT) sodium salt was from Sigma-Aldrich. CytoTox96 LDH-release kit was from Promega. The Cy3-labelled plasmid DNA (MIR 7904) was obtained from Mirus Bio, LLC.

**Full-length cDNAs and expression constructs.** The pCMV-sport6 mammalian expression plasmids for human full-length AIM2 (IMAGE:30367168), IFIX (IMAGE:5212268) and MND A (IMAGE:5223430) were obtained from Open Biosystems and their sequences were verified by automated DNA sequencing. pcDNA3-IFI16 plasmid was a gift from T. Ouchi. pcDNA3-AIM2-T7 and pcDNA3-AIM2-ΔPYD-T7 were generated by cloning PCR-generated full-length AIM2 or AIM2-ΔPYD cDNAs, respectively, in a modified pcDNA3-T7 plasmid in frame with a C-terminal T7 tag. For stable expression of a C-terminal EGFP-tagged human or mouse AIM2 in 293T cells or in mouse *NALP3*<sup>-/-</sup> immortalized macrophages, the AIM2 cDNAs were fused in-frame with EGFP cDNA in the pEGFP-N1 plasmid. The AIM2-EGFP fusion cDNAs were then excised from the pEGFP-N1 plasmids and subcloned into the pMSCVpuro retroviral transfer vector to generate the pMSCVpuro-hAIM2-EGFP-N1 and pMSCVpuro-mAIM2-EGFP-N1 vectors. For expression of AIM2 or AIM2-ΔPYD with a His6 or GST tag in bacteria, PCR-generated full-length AIM2 or AIM2-ΔPYD cDNAs were inserted into the bacterial expression vector pET-21a (+) in-frame with a C-terminal His6 or GST tag. Bacterial expression constructs for GST-tagged ASC, ASC-PYD or ASC-CARD were described previously<sup>17</sup>. The mammalian expression plasmids for full-length human pyrin were described previously (pcDNA-pyrin-myc-HisB)<sup>9</sup>.

**Generation of stable cell lines.** 293T-caspase-1, 293T-caspase-1-ASC and 293T-ASC-EGFP-N1 cell lines have been described previously<sup>4,8,9</sup>. Stable 293T-caspase-1-pro-IL1β or 293T-caspase-1-ASC-pro-IL1β were generated by transfecting 293T-caspase-1 or 293T-caspase-1-ASC, respectively, with dual expression plasmid pMSCVgfp-pro-IL1β. Stable cell lines were obtained after multiple cell sorting over a period of one month by flow cytometry as described before<sup>8</sup>. Stable 293T-caspase-1-ASC-AIM2 cells were generated by transfecting 293T-caspase-1-ASC with dual expression plasmid pRSC-GFP-AIM2 followed by sorting as above. Stable 293T-AIM2-ASC-EGFP-N1 cell line was generated by co-transfecting 293T cells with pMSCVpuro-ASC-EGFP-N1<sup>8</sup> and pcDNA3-AIM2-T7 plasmids at a 1:4 ratio, followed by cell sorting as above and selection in Geneticin (G418, GIBCO) and puromycin-containing medium. Stable 293T-AIM2-EGFP-N1 cells, expressing the human AIM2-EGFP fusion protein, were generated by transfecting 293T cells with pMSCVpuro-hAIM2-EGFP-N1 expression plasmid followed by cell sorting as described previously and selection with puromycin-containing medium. Stable *NALP3*<sup>-/-</sup>-AIM2-EGFP-N1, which express the mouse AIM2-EGFP fusion protein, were generated by infecting the *NALP3*<sup>-/-</sup> immortalized macrophages with pMSCVpuro-mAIM2-EGFP-N1 retroviral vector followed by cell sorting as above and selection with puromycin containing medium.

**siRNA knockdown.** Human AIM2-targeting siRNA oligonucleotides (AIM2 Stealth Select RNAi, catalogue number 1299003) were obtained from

Invitrogen. The sequences of each siRNA oligonucleotide in this pool are as follows: *hAIM2*-HSS114049, 5'-UAUGGUGCUAUGAACUCCAGAUAGUC-3'; *hAIM2*-HSS114050, 5'-UUUCAGCUUGACUAGUGGCCUUUGG-3'; *hAIM2*-HSS114051, 5'-UUCUCUGAUGAUCCUGGGCC-3'. Mouse *Aim2*-targeting siRNA oligonucleotides (ON-TARGETplus SMARTpool) were obtained from Dharmacon. The sequences of each siRNA oligonucleotide in this pool are as follows: *mAim2* J044968-09, 5'-ACAUGACACUGAGGGUUAU-3'; *mAim2* J044968-10, 5'-UGUCUAAGGCUUGGGUAU-3'; *mAim2* J044968-11, 5'-CUACCUGAGGAUAGCAUUU-3'; *mAim2* J044968-12, 5'-AGUACUAGG-AAAUCAGUGA-3'. Non-targeting control siRNA oligonucleotides (AllStars Neg. Control siRNA, catalogue number 1027281) were obtained from Qiagen. For AIM2 knockdown experiments in the human THP-1 cell line, cells were seeded in 35-mm 6-well plates at a density of 2 × 10<sup>6</sup> cells per well in the presence of 1 μM PMA and allowed to attach to culture dishes for 3 h. Cells were then washed and treated with LPS (2 μg ml<sup>-1</sup>) for 16 h to induce differentiation. The differentiated cells were then transfected with siRNA oligonucleotides (50 nm) for 24 h using Lipofectamine 2000 (3.5 μl ml<sup>-1</sup>) in 2 ml of SF OPTI-MEM I medium. The next day, cells were washed with SF OPTI-MEM I medium and then transfected with poly(dA:dT) for 6 h with Lipofectamine 2000 in 2 ml of SF OPTI-MEM I. The culture supernatants and cells were separated and then processed for immunoblot analysis. For AIM2 knockdown experiments in the immortalized mouse macrophages, cells were seeded in 6-well plates at a density of 2 × 10<sup>6</sup> cells per well in SF OPTI-MEM I medium and allowed to attach for 2 h. The cells were then transfected with siRNA oligonucleotides (50 nm) for 24 h using Lipofectamine 2000 in 2 ml of SF OPTI-MEM I medium. The next day, cells were washed with SF OPTI-MEM I medium and then transfected with poly(dA:dT) (1 μg ml<sup>-1</sup>) for 6 h with Lipofectamine 2000 in 2 ml of SF OPTI-MEM I. The culture supernatants and cells were separated and then processed for immunoblot analysis.

**Immunoblot analysis.** Cell culture supernatants were precipitated by the addition of an equal volume of methanol and 0.25 volumes of chloroform as described previously<sup>7</sup>. The supernatant/methanol/chloroform mixtures were vortexed and then centrifuged for 10 min at 20,000g. The upper phase was discarded and 500 μl methanol was added to the interphase. This mixture was centrifuged for 10 min at 20,000g and the protein pellet was dried at 55 °C, resuspended in Laemmli buffer and boiled for 5 min at 99 °C. Samples were separated by 12.5% SDS-PAGE and were transferred onto nitrocellulose membranes. Blots were probed with rabbit polyclonal antibody to the human caspase-1 p30, rabbit polyclonal antibody to human cleaved IL1β (Asp116; Cell Signaling) or rat monoclonal anti-mouse caspase-1 p20 antibody (a gift from J. Yuan). Total cell lysates were mixed with SDS sample buffer, fractionated on 12.5% SDS-PAGE and then immunoblotted as described previously.

**Immunoprecipitation and in vitro GST pull-down assays.** THP-1 cells (10<sup>7</sup> cells) were treated with IFN-γ (75 ng ml<sup>-1</sup>) for 24 h and then lysed in 250 μl buffer A (20 mM HEPES, pH 7.5, 50 mM KCl, 150 mM NaCl, 1.5 mM MgCl<sub>2</sub>, 1 mM EGTA, 0.5% NP-40, 1 mM DTT, 0.2 mM PMSF, 2 μg ml<sup>-1</sup> aprotinin, 5 mM NaF, 1 mM Na<sub>3</sub>VO<sub>4</sub>). The cell lysate was centrifuged at 20,000g to remove cell debris. The supernatant was diluted twice with a hypotonic lysis buffer (10 mM HEPES, pH 7.8, 10 mM KCl, 2 mM MgCl<sub>2</sub>, 0.1 mM EDTA and protease inhibitor) and then immunoprecipitated with anti-human ASC or anti-human Omi mouse monoclonal antibodies at 4 °C. The protein-antibody complexes were then precipitated with protein G-sepharose bead, and the bead-bound complexes and lysates were then fractionated by SDS-PAGE followed by immunoblotting with anti-AIM2 (3B10) or anti-IFI16 mouse monoclonal antibodies, or anti-ASC, anti-Omi, anti-MND A or anti-IFI16 rabbit polyclonal antibodies. *In vitro* pull-down assays were performed with recombinant bacterially expressed GST-tagged full-length human ASC, ASC-PYD or ASC-CARD and C-terminal His6-tagged human AIM2 or AIM2-ΔPYD. The GST-tagged proteins were isolated from bacterial lysates by glutathione-affinity purification on glutathione-agarose beads. The bead-bound proteins were then incubated for 2 h at 4 °C with bacterially produced His6-tagged AIM2 or AIM2-ΔPYD proteins. After incubation, the complexes were fractionated by SDS-PAGE and detected by immunoblotting with anti-His6-HRP antibody. Similarly GST-tagged AIM2 or AIM2-ΔPYD was allowed to interact with bacterially produced ASC and the bound ASC was assayed by immunoblotting with anti-ASC monoclonal antibody.

**Confocal microscopy.** The 293T-ASC-EGFP-N1 or 293T-AIM2-ASC-EGFP-N1 cells, or *NALP3*<sup>-/-</sup> macrophages, were seeded on 35 mm cover glass-bottom culture dishes and allowed to attach for 24 h. The next day cells were transfected with poly(dA:dT) (1 μg per dish) for 2–3 h using Lipofectamine 2000 and then stained with Hoechst 33342 for 30 min. Cells were then observed using a Zeiss LSM 510 Meta confocal microscope. In the experiments using the Cy3-labelled plasmid DNA, 293T-AIM2-EGFP-N1 cells or *NALP3*<sup>-/-</sup>-AIM2-EGFP-N1 macrophages were seeded on 35 mm cover glass-bottom culture dishes and allowed to attach for 24 h. The next day cells were transfected with Cy3-labelled

plasmid DNA (0.5 µg per dish) for 2–3 h using Lipofectamine 2000 and then stained with Hoechst 33342 for 30 min. Cells were then observed using a Zeiss LSM 510 Meta confocal microscope. The EGFP (green) was excited with the 488 nm argon laser. The nuclear Hoechst 33342 stain (blue) was excited with the 405 nm diode laser. The Cy3 (red) was excited with the 543 nm He/Ne Laser.

**In vitro caspase-1 activation and IL1 $\beta$  cleavage assays.** Lysates containing PYD–caspase-1 and pro-IL1 $\beta$  were prepared from 293T-PYD–caspase-1 or 293T-pro-IL1 $\beta$  stable cell lines in CHAPS buffer (20 mM HEPES-KOH, pH 7.5, 5 mM MgCl<sub>2</sub>, 0.5 mM EGTA, 0.1 mM PMSF, 0.1% CHAPS). GST–AIM2 or ASC pyroptosomes were produced in bacteria and purified by standard procedures<sup>18</sup>. Purified GST–AIM2 or ASC pyroptosomes were incubated with PYD–caspase-1-containing lysates together with pro-IL1 $\beta$  in CHAPS buffer for 60 min at 37 °C. The reaction mixtures were then fractionated by SDS–PAGE and analysed by western blotting with anti-caspase-1 p30 and anti-IL1 $\beta$  antibodies. In some experiments, *in vitro* translated <sup>35</sup>S-labelled and nuclease-treated PYD–caspase-9 (wild type or C287A) was incubated with purified GST–AIM2 or ASC pyroptosomes in the presence or absence of poly(dA:dT) for 60 min at 37 °C and the reaction products then analysed by SDS–PAGE followed by autoradiography.

**Pyroptosome isolation from poly(dA:dT) transfected macrophages.** *NALP3*<sup>−/−</sup> macrophages in 10 cm culture dishes were transfected with 6 µg of poly(dA:dT) or vehicle (H<sub>2</sub>O) using Lipofectamine 2000 in SF OPTI-MEM I for 4 h. The cells were then pelleted by centrifugation and then lysed in buffer A

(20 mM HEPES-KOH, pH 7.5, 10 mM KCl, 1.5 mM MgCl<sub>2</sub>, 1 mM EDTA, 1 mM EGTA, 320 mM sucrose, 0.1 mM PMSF). ASC pyroptosomes were then isolated from the lysates and cross-linked with DSS as described in detail recently<sup>18</sup>.

**Purification and cross-linking of AIM2.** Full-length AIM2 with a C-terminal His6 tag was expressed in bacteria and then purified on Talon beads (Clontech) according to standard procedures. Purified AIM2 (2 µg) was incubated with or without a 64-mer synthetic dsDNA (0.5 and 1 µg) for 15 min at room temperature in 50 µl PBS. The reaction mixtures were then cross-linked with 1 µl of 1% glutaraldehyde, which covalently cross-links lysine residues in close proximity to each other in oligomeric proteins. Glutaraldehyde does not cross-link proteins to DNA. The reactions were terminated by addition of 2 µl of 2 M Tris-HCl, pH 8.0. Cross-linked AIM2 was solubilized by addition of an equal volume of 2× Laemmli SDS sample buffer and fractionated on a 12.5% SDS–polyacrylamide gel.

**Statistics.** All values are expressed as the mean and s.d. of individual samples. Samples were analysed using the Student's *t*-test.

17. Srinivasula, S. M. *et al.* The PYRIN-CARD protein ASC is an activating adaptor for caspase-1. *J. Biol. Chem.* **277**, 21119–21122 (2002).
18. Fernandes-Alnemri, T. & Alnemri, E. S. Chapter thirteen assembly, purification, and assay of the activity of the ASC pyroptosome. *Methods Enzymol.* **442**, 251–270 (2008).

## LETTERS

# AIM2 recognizes cytosolic dsDNA and forms a caspase-1-activating inflammasome with ASC

Veit Hornung<sup>1,2</sup>, Andrea Ablasser<sup>1,2</sup>, Marie Charrel-Dennis<sup>1</sup>, Franz Bauernfeind<sup>1,2</sup>, Gabor Horvath<sup>1</sup>, Daniel R. Caffrey<sup>3</sup>, Eicke Latz<sup>1\*</sup> & Katherine A. Fitzgerald<sup>1\*</sup>

The innate immune system senses nucleic acids by germline-encoded pattern recognition receptors. RNA is sensed by Toll-like receptor members TLR3, TLR7 and TLR8, or by the RNA helicases RIG-I (also known as DDX58) and MDA-5 (IFIH1)<sup>1</sup>. Little is known about sensors for cytoplasmic DNA that trigger antiviral and/or inflammatory responses<sup>2–6</sup>. The best characterized of these responses involves activation of the TANK-binding kinase (TBK1)–interferon regulatory factor 3 (IRF3) signalling axis to trigger transcriptional induction of type I interferon genes<sup>2,3</sup>. A second, less well-defined pathway leads to the activation of an ‘inflammasome’ that, via caspase-1, controls the catalytic cleavage of the pro-forms of the cytokines IL1 $\beta$  and IL18 (refs 6, 7). Using mouse and human cells, here we identify the PYHIN (pyrin and HIN domain-containing protein)<sup>8</sup> family member absent in melanoma 2 (AIM2) as a receptor for cytosolic DNA, which regulates caspase-1. The HIN200 domain of AIM2 binds to DNA, whereas the pyrin domain (but not that of the other PYHIN family members) associates with the adaptor molecule ASC (apoptosis-associated speck-like protein containing a caspase activation and recruitment domain) to activate both NF- $\kappa$ B and caspase-1. Knockdown of *Aim2* abrogates caspase-1 activation in response to cytoplasmic double-stranded DNA and the double-stranded DNA vaccinia virus. Collectively, these observations identify AIM2 as a new receptor for cytoplasmic DNA, which forms an inflammasome with the ligand and ASC to activate caspase-1.

Our current understanding of the mechanisms sensing cytoplasmic DNA is limited<sup>9</sup>. A candidate receptor called DAI (DNA-dependent activator of interferon (IFN)-regulatory factors) has been implicated in the DNA-induced type I IFN pathway<sup>4</sup>. The NLR family member NLRP3 has also been shown to activate caspase-1 in response to internalized adenoviral DNA<sup>6</sup>. Caspase-1 activation in response to transfected bacterial, viral, mammalian or synthetic DNA, however, does not involve NLRP3, although the adaptor molecule ASC is required<sup>6,7</sup>.

We proposed that an upstream activator of this double-stranded DNA (dsDNA)-activated ASC pathway would contain a pyrin domain (PYD) for homotypic interaction with ASC, and at least one other domain for direct binding to DNA or for association with an upstream receptor. In addition to NLRP3 (ref. 10), NLRP6 (ref. 11) and NLRP12 (ref. 12) have previously been shown to associate with ASC. Although ASC-deficient macrophages failed to activate caspase-1 and trigger IL1 $\beta$  release in response to poly(dA:dT) • poly(dA:dT) (hereafter referred to as poly(dA:dT))<sup>2</sup>, macrophages lacking NLRP3, NLRP6 and NLRP12 responded normally (Fig. 1a). Surprisingly, we found that macrophages lacking ASC produced higher levels of IFN $\beta$  and IL6 in response to poly(dA:dT), which was not observed in cells

lacking NLRP3, the IL1 receptor, or to a lesser extent caspase-1 (Supplementary Fig. 1a–d). Poly(dA:dT)-induced cell death also occurred in an ASC-dependent manner (Supplementary Fig. 1e, f). We speculate that the increased cytokine production in ASC-deficient cells relates to their resistance to poly(dA:dT)-induced cell death. In addition to poly(dA:dT), dsDNA from natural sources activated caspase-1 cleavage (Supplementary Fig. 2a, b). In contrast, a small immunostimulatory oligonucleotide<sup>3</sup>, long single-stranded DNA (ssDNA; poly(dI)), transfected dsRNA or the ssRNA virus Sendai virus failed to trigger this response in *Nlrp3*-deficient macrophages (Supplementary Fig. 2c).

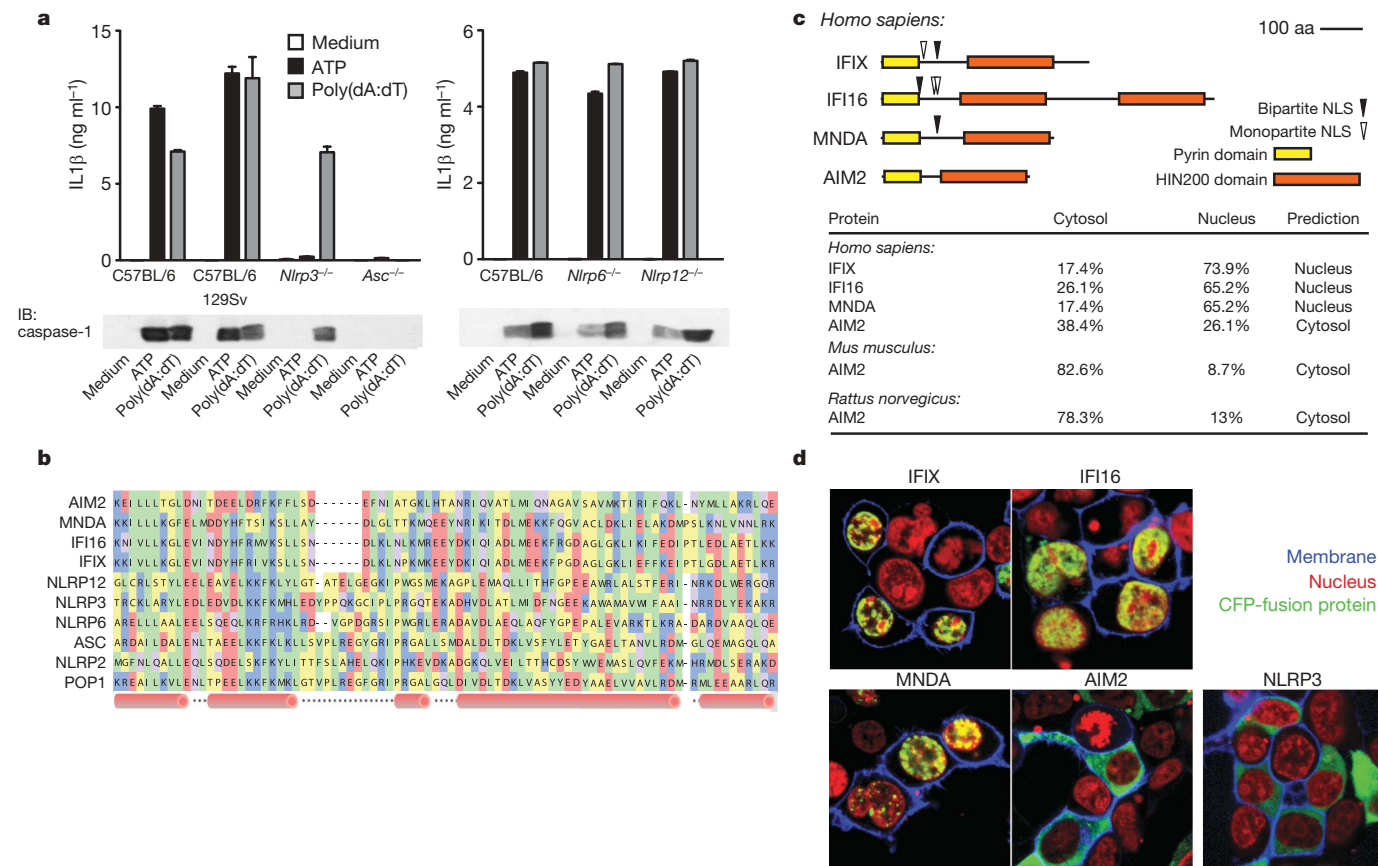
Searching the PFAM database<sup>13</sup> we identified several PYD-domain containing proteins, which also contained a HIN200 domain, previously shown to bind DNA<sup>14</sup>. In humans, the HIN200 family consists of four members<sup>15</sup>: IFIX (also known as PYHIN1)<sup>16</sup>, IFI16 (ref. 17), MNDA<sup>18</sup> and AIM2 (ref. 19). A multiple-sequence alignment of PYD domains from these proteins with PYD domains from some of the NLR proteins is shown in Fig. 1b. Sequence analysis of IFIX, IFI16 and MNDA predicted their nuclear localization, in contrast to AIM2, which was predicted to be cytosolic (Fig. 1c, bottom panel). Consistent with these predictions, fluorescent protein chimaeras of IFIX, IFI16 and MNDA localized to the nucleus, whereas AIM2 was almost exclusively cytoplasmic (Fig. 1d).

To study the possibility that these PYHIN proteins associated with ASC, we generated carboxy-terminally tagged PYD–CFP (cyan fluorescent protein)-domain fusions (which lacked the putative nuclear localization sequences identified above). Indeed, all of the PYD–CFP fusions were localized to the cytoplasm (Supplementary Fig. 3). To test whether induced clustering of the PYD–CFP fusions led to association with ASC–YFP (yellow fluorescent protein), we used a human embryonic kidney 293 cell line that stably expressed ASC–YFP at low enough levels to be polydispersed throughout the cytoplasm (Fig. 2a, mock). Indeed, overexpression of the NLRP3–CFP-tagged PYD domain led to the formation of large cytosolic aggregates, which co-aggregated with ASC–YFP (Fig. 2a). Notably, in most transfected cells, extensive intracellular co-localization with ASC–YFP was observed with a complete loss of the cytoplasmic distribution of ASC–YFP<sup>10,11,20</sup>. Of all the PYHIN–PYD proteins tested, only AIM2–PYD led to complex formation with ASC (Fig. 2a, b and Supplementary Fig. 4). Similar results were obtained with full-length AIM2–CFP but not with full-length IFIX, IFI16 or MNDA, which were all localized to the nucleus (Fig. 2a, b and Supplementary Fig. 5a, b). Additionally, only AIM2–PYD and NLRP3–PYD were found to bind haemagglutinin (HA)-tagged ASC in co-immunoprecipitation studies (Fig. 2c). Furthermore, endogenous ASC associated with endogenous AIM2, but not with IFI16 in primed THP-1 cells (Fig. 2d).

<sup>1</sup>Division of Infectious Diseases and Immunology, Department of Medicine, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA. <sup>2</sup>Institute of Clinical Chemistry and Pharmacology, Universitätsklinikum Bonn 53127, Germany. <sup>3</sup>Pfizer, 620 Memorial Drive, Cambridge, Massachusetts 02139, USA.

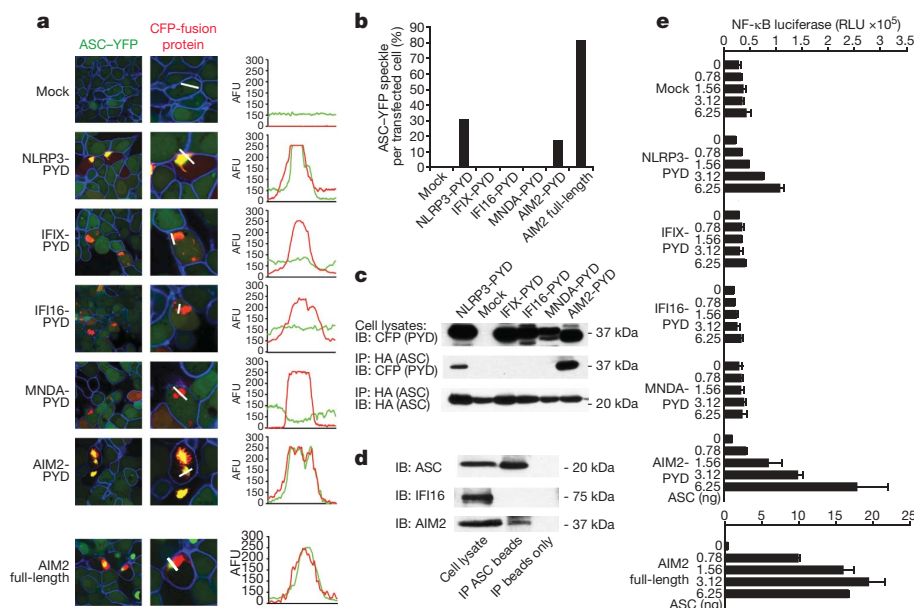
\*These authors contributed equally to this work.





**Figure 1 | Poly(dA:dT)-induced inflammasome activation.** **a**, LPS-primed macrophages from wild type or inflammasome-deficient mice were stimulated as indicated and supernatants were examined for IL1 $\beta$  by ELISA or for cleaved caspase-1 by immunoblot. Error bars represent s.d. **b**, A multiple-sequence alignment of human PYHIN and select NLR PYD domains. **c**, Domain structures of human PYHINs, with predicted nuclear

localization signals and subcellular localizations (bottom panel). aa, amino acids; NLS, nuclear localization sequence. **d**, Subcellular localization of CFP-tagged IFIX, IFI16, MNDA, AIM2 or NLRP3 (all green) in 293T cells. Fluorescent cholera-toxin-stained membranes (blue) and DRAQ5-stained nuclei (red). Original magnification,  $\times 150$ . Data from one experiment of three is shown (**a**, **d**).



**Figure 2 | AIM2 interacts with ASC.** **a**, ASC-YFP-expressing cells (green) were transfected as indicated and imaged by confocal microscopy. Original magnification,  $\times 150$ . AFU, arbitrary fluorescent units. **b**, The fluorescence intensities of green (ASC-YFP) and red (PYD-CFP) channels were quantified along the white lines in **a** and the percentage of ASC-YFP speckles was calculated. **c**, 293T cells transfected with HA-ASC and CFP-tagged constructs as in **a** were immunoprecipitated (IP) with an anti-HA antibody

and immunoblotted (IB) as indicated. **d**, ASC was immunoprecipitated from Sendai-virus-primed THP-1 cells, and ASC, IFI16 and AIM2 were examined by immunoblotting. The band above the AIM2 band in the ASC immunoprecipitation corresponds to the heavy chain of the anti-ASC antibody. **e**, NF- $\kappa$ B luciferase reporter gene activity was measured on transfection with the indicated plasmids. Data are representative of one experiment out of three (**a-d**) or out of two (**e**). Error bars, s.d.

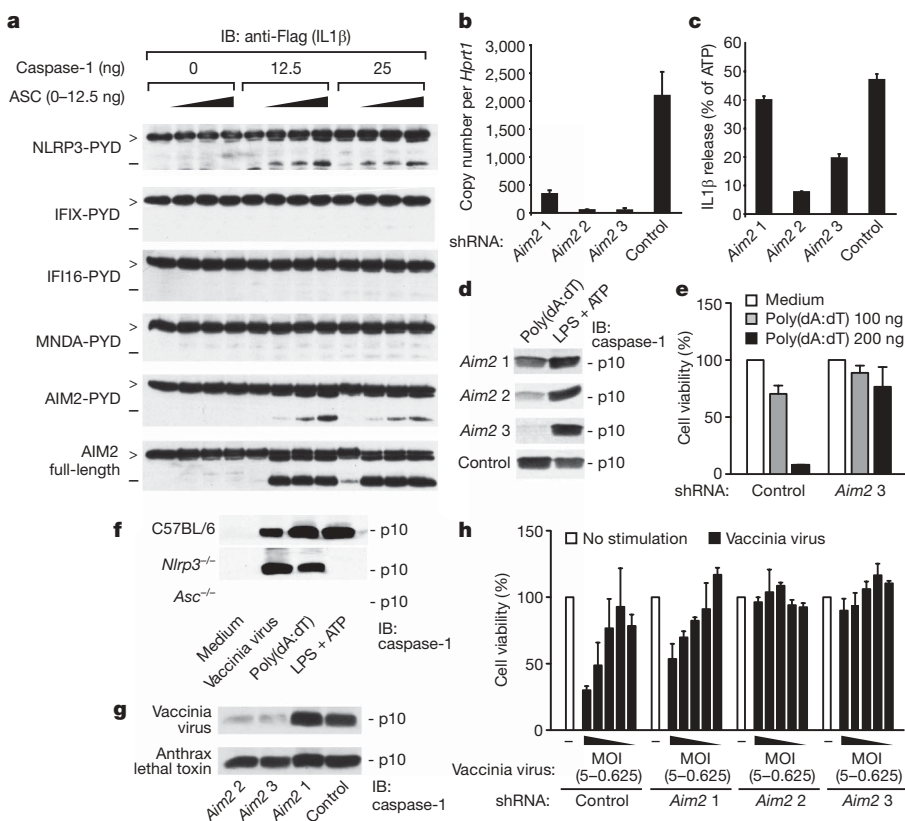
To examine the functional relevance of AIM2–ASC complex formation, we examined NF- $\kappa$ B reporter gene activity in cells overexpressing the PYHIN-PYD proteins in the presence of ASC. Only NLRP3-PYD and AIM2-PYD led to potent NF- $\kappa$ B activation (Fig. 2e). The effect of full-length AIM2 was even more marked (Fig. 2e, bottom panel). The full-length versions of IFIX, IFI16 and MNDA failed to activate NF- $\kappa$ B (Supplementary Fig. 5c). ASC was absolutely required, because no substantial NF- $\kappa$ B reporter activity was observed in cells not transfected with ASC. No substantial activation of the IFN $\beta$  promoter reporter gene was observed with any of the PYHIN family members (Supplementary Fig. 6).

We next examined whether the AIM2–ASC complex could lead to the formation of a functional inflammasome complex and caspase-1-dependent maturation of pro-IL1 $\beta$ . We used a transient transfection assay overexpressing the respective proteins of interest in the presence of ASC, caspase-1 and Flag-tagged pro-IL1 $\beta$  in 293T cells and monitored the cleavage of pro-IL1 $\beta$  by immunoblotting. Among the PYD proteins tested, only NLRP3-PYD and AIM2-PYD induced maturation of pro-IL1 $\beta$ , when ASC and caspase-1 were co-expressed (Fig. 3a). Full-length AIM2 was even more potent than AIM2-PYD (Fig. 3a, lower panel). Neither the PYD domain nor the full-length versions of IFIX, IFI16 or MNDA induced IL1 $\beta$  cleavage (Supplementary Fig. 7).

To study the role of AIM2 in cells with a functional poly(dA:dT)-triggered or dsDNA-virus-induced inflammasome complex, we used lentiviruses encoding short hairpin RNAs (shRNAs) to knock down *Aim2* in immortalized murine macrophage cell lines (B6-MCLs or N3-KO-MCLs)<sup>7</sup>. AIM2 was expressed constitutively both in primary macrophages and in B6-MCLs, and was further induced by poly(dA:dT) or Sendai virus (Supplementary Fig. 8). Three different shRNAs were tested, of which two (*Aim2* shRNA2 and shRNA3) resulted in a strong reduction of *Aim2* expression (Fig. 3b). Knocking down *Aim2*, but not an unrelated gene, resulted in a strong attenuation of both poly(dA:dT)-mediated IL1 $\beta$  release (Fig. 3c) and caspase-1 cleavage (Fig. 3d). Targeting human *AIM2* in THP-1 cells using short interfering RNA (siRNA) corroborated these findings

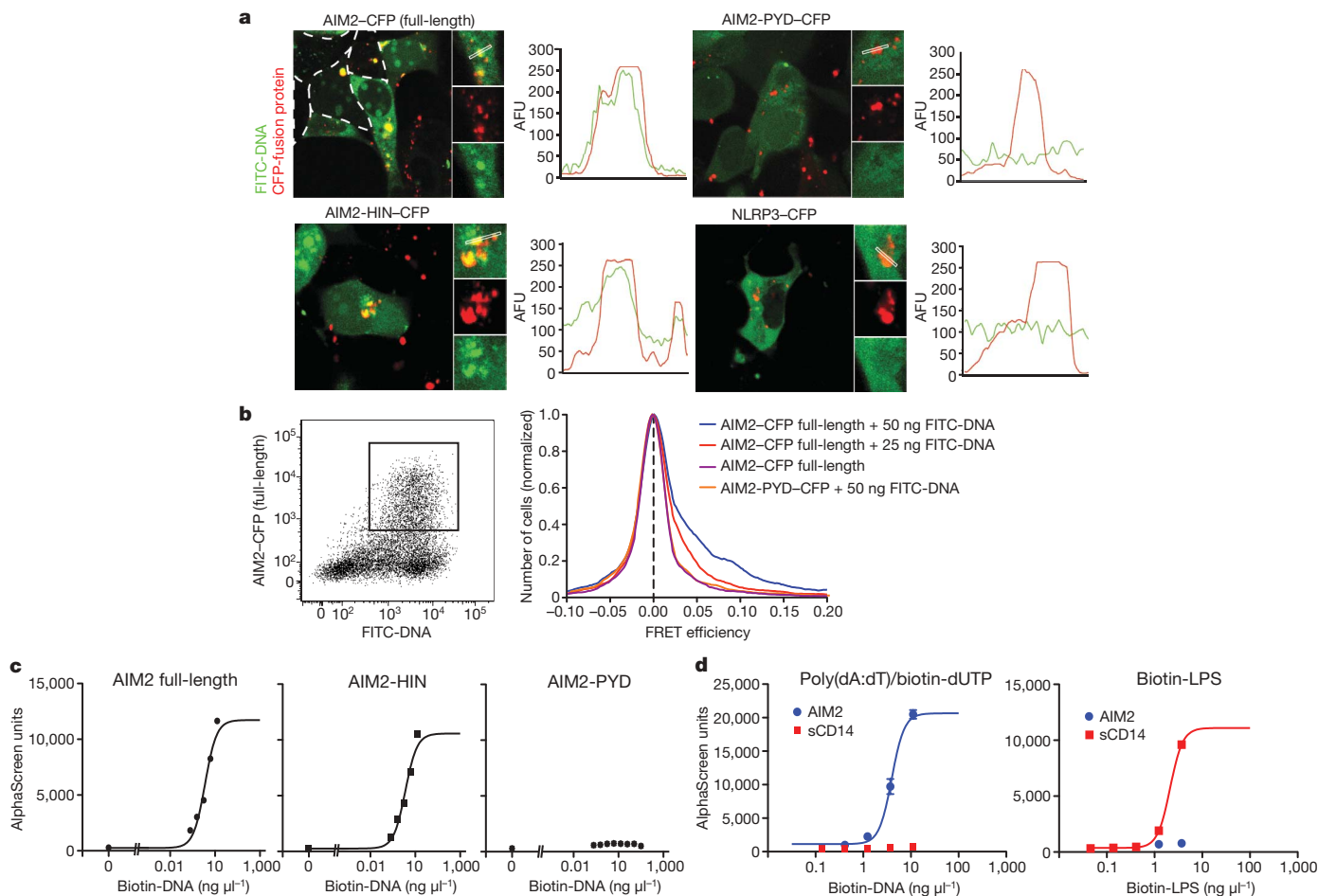
(Supplementary Fig. 8d, e). Moreover and consistent with what we had seen in ASC-deficient macrophages (Supplementary Fig. 1), knocking down *Aim2* resulted in a marked enhancement of poly(dA:dT)-mediated type I IFN induction (Supplementary Fig. 8b). This effect was specific because the IFN $\beta$  response to Sendai virus was unaffected (Supplementary Fig. 8c). Furthermore, in agreement with the results obtained in ASC-deficient macrophages, macrophages that were targeted with *Aim2* shRNAs were resistant to poly(dA:dT)-triggered cell death (Fig. 3e). We also examined the role of AIM2 in the recognition of the dsDNA vaccinia virus. Similar to what we had observed with transfected poly(dA:dT), vaccinia-virus-induced caspase-1 cleavage occurred in an ASC-dependent but an NLRP3-independent manner (Fig. 3f). This effect was also dependent on AIM2, because shRNA-mediated knockdown of *Aim2* impaired vaccinia-virus-induced caspase-1 cleavage but not that induced by anthrax lethal toxin (Fig. 3g). Knockdown of a control protein did not affect caspase-1 cleavage after vaccinia virus infection. Vaccinia-virus-triggered cell death was also strongly reduced in *Aim2*-shRNA-targeted macrophages, but not in control macrophages (Fig. 3h). Altogether, these results indicated that AIM2 controlled inflammasome activation and cell death in response to dsDNA and the dsDNA vaccinia virus.

To determine whether AIM2 could be involved in the recognition of dsDNA directly, we generated fluorescein-labelled poly(dA:dT), (FITC-DNA), and co-transfected FITC-DNA together with CFP-tagged versions of full-length AIM2, AIM2-HIN, AIM2-PYD or full-length NLRP3. Whereas cells expressing NLRP3 or AIM2-PYD showed no co-localization of the respective proteins with FITC-DNA, full-length AIM2 and AIM2-HIN showed extensive co-localization with FITC-DNA and led to the formation of DNA and protein aggregates in the cytosol (Fig. 4a). We used single cell flow cytometry fluorescence resonance energy transfer (FRET) measurements to quantify these interactions (Fig. 4b)<sup>21</sup>. A dose-dependent increase in FRET between full-length AIM2 and FITC-DNA was seen, whereas AIM2-PYD did not lead to measurable FRET. Other proteins such as NLRP3 and IFI16 did not show any FRET (data not shown).



**Figure 3 | AIM2 is required for poly(dA:dT)- and vaccinia-virus-triggered inflammasome activation.**

**a**, 293T cells were transfected as indicated and the cell lysates were immunoblotted for pro-IL1 $\beta$  (>) and cleaved IL1 $\beta$  (<). **b**, B6-MCLs were transduced with lentiviral vectors encoding shRNAs as indicated, and *Aim2* and *Hprt1* were measured by quantitative rtPCR. **c**, LPS-primed cells as in **b** were stimulated as indicated for 6 h and IL1 $\beta$  in the supernatants was measured by ELISA. The poly(dA:dT)-triggered IL1 $\beta$  release was normalized to the ATP-induced IL1 $\beta$  levels. Absolute values for ATP-triggered IL1 $\beta$  release were 1,790, 2,078, 2,676 and 1,119 pg ml<sup>-1</sup> for *Aim2* shRNA1, shRNA2, shRNA3 and control, respectively. **d**, Macrophages as in **b** were treated as indicated, and the cleavage of caspase-1 was measured by immunoblotting after 6 h. **e**, Macrophages transduced with shRNA were transfected as indicated and cells were counted 24 h later. **f**, Macrophages from the indicated strains were treated as indicated and cleavage of caspase-1 was measured after 6 h. **g**, *Nlrp3*-deficient shRNA-expressing macrophages as in **b** were infected as indicated and assessed for cleavage of caspase-1 after 6 h. **h**, Cells as in **b** were infected as indicated and cell survival was measured by calcein AM staining 24 h later. MOI, multiplicity of infection. One representative experiment out of three (**a**, **d–h**), four (**c**) or five (**b**) is depicted. Error bars represent s.d.



**Figure 4 | AIM2 binds dsDNA via its HIN domain.** **a**, 293T cells were transfected with CFP-tagged AIM2 or NLRP3 (both red) as indicated, together with FITC-DNA (green), and imaged by confocal microscopy. Fluorescence intensities of the green (FITC-DNA) and red (CFP-fusion protein) channels were quantified along selected lines. Original magnification,  $\times 150$ . **b**, 293T cells were transfected as indicated with FITC-DNA or unlabelled DNA. Cells were analysed by flow cytometry after 24 h

and CFP- and FITC-positive cells were gated (left panel) and analysed for FRET efficiency on a cell-by-cell basis. Calculated FRET efficiency histograms are shown (right panel). **c**, **d**, AlphaScreen assessment of AIM2 full-length, AIM2-HIN or AIM2-PYD binding to poly(dA:dT)/biotin-dUTP (biotin-DNA) (**c**) and AIM2 or soluble CD14 (sCD14) binding to biotin-DNA or biotin-LPS (**d**). Representative data from two (**d**) or three (**a–c**) independent experiments are shown.

Furthermore, binding studies using purified AIM2, AIM2-HIN and AIM2-PYD with biotinylated poly(dA:dT) (biotin-DNA) showed that AIM2 directly interacted with poly(dA:dT) with high affinity; only full-length AIM2 or the AIM2-HIN was able to bind biotin-DNA (Fig. 4c). Binding of poly(dA:dT) to AIM2 was specific, because AIM2 did not bind biotin-lipopolysaccharide (LPS), which bound to soluble CD14 under similar assay conditions (Fig. 4d).

Collectively, these data identify AIM2 as a receptor for cytosolic dsDNA, which forms a new inflammasome complex with ASC to activate caspase-1-mediated processing of IL1 $\beta$ . Our data also indicate that the activation of the AIM2 inflammasome is important in innate immunity to vaccinia virus. Because bacterial pathogens such as *Francisella tularensis*<sup>22</sup> and aberrant host DNA in pathological autoimmunity<sup>23</sup> also trigger the IL1 $\beta$  pathway, it will be important to define the role of AIM2 in these responses. Further characterization of the AIM2 inflammasome as a sensor of microbial, as well as host DNA, may therefore enable the rational design of new therapies and treatments for infectious as well as autoimmune diseases.

## METHODS SUMMARY

**Reagents and mice.** All complementary DNAs were cloned by PCR from cDNA into pEFBOS-C-term-CFP and subcloned into pEFBOS-C-term-Flag/His. Biotinylated and FITC-labelled poly(dA:dT) were made by adding biotin-dUTP or FITC-dUTP (Fermentas) at a molar ratio of 1:8 to dTTP in the enzymatic synthesis of poly(dA:dT) as described<sup>24</sup>. Vaccinia virus (Western Reserve

strain) was from K. Rock. The anti-human AIM2 antibody (3B10) was from R. Johnstone. *Nlrp3*<sup>-/-</sup> and *Pycard*<sup>-/-</sup> mice were as previously described<sup>25</sup>. Both strains, as well as *Nlrp6*<sup>-/-</sup> and *Nlrp12*<sup>-/-</sup> mice, were from Millennium Pharmaceuticals. Caspase-1-deficient mice were from R. Flavell. *Il1r1*<sup>-/-</sup> mice were from Jackson Laboratories.

Bone-marrow-derived macrophages were stimulated as indicated. Poly(dA:dT) DNA and all other DNAs were transfected using Lipofectamine 2000 at a concentration of  $1 \mu\text{g ml}^{-1}$ . Cell culture supernatants were assayed for IL1 $\beta$  using ELISA kits from BD Biosciences. Confocal microscopy was performed on a Leica SP2 AOBs confocal laser scanning microscope. FRET efficiencies were calculated on a cell-by-cell basis<sup>21</sup> and histograms were plotted with GraphPad Prism 5.01 (GraphPad Software). Immunoblot analysis was conducted as previously described<sup>7</sup>. Quantitative real-time PCR (rtPCR) analysis was performed as previously described<sup>26</sup>. Lentiviral shRNAs targeting *Aim2* were obtained from OpenBiosystems and shRNA silencing was carried out as described ([http://www.broad.mit.edu/genome\\_bio/trc/publicProtocols.html](http://www.broad.mit.edu/genome_bio/trc/publicProtocols.html)). The AlphaScreen (amplified luminescent-proximity homogeneous assay) was set up as an association assay and read with the Envision HT microplate reader (Perkin Elmer). Reporter assays for NF- $\kappa$ B or IFN luciferase reporters were carried out as previously described<sup>26</sup>.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 10 November; accepted 15 December 2008.

Published online 21 January 2009.

- Meylan, E., Tschopp, J. & Karin, M. Intracellular pattern recognition receptors in the host response. *Nature* **442**, 39–44 (2006).



2. Ishii, K. J. *et al.* A Toll-like receptor-independent antiviral response induced by double-stranded B-form DNA. *Nature Immunol.* **7**, 40–48 (2006).
3. Stetson, D. B. & Medzhitov, R. Recognition of cytosolic DNA activates an IRF3-dependent innate immune response. *Immunity* **24**, 93–103 (2006).
4. Takaoka, A. *et al.* DAI (DLM-1/ZBP1) is a cytosolic DNA sensor and an activator of innate immune response. *Nature* **448**, 501–505 (2007).
5. Ishii, K. J. *et al.* TANK-binding kinase-1 delineates innate and adaptive immune responses to DNA vaccines. *Nature* **451**, 725–729 (2008).
6. Muruve, D. A. *et al.* The inflammasome recognizes cytosolic microbial and host DNA and triggers an innate immune response. *Nature* **452**, 103–107 (2008).
7. Hornung, V. *et al.* Silica crystals and aluminum salts activate the NALP3 inflammasome through phagosomal destabilization. *Nature Immunol.* **9**, 847–856 (2008).
8. Ludlow, L. E., Johnstone, R. W. & Clarke, C. J. The HIN-200 family: more than interferon-inducible genes? *Exp. Cell Res.* **308**, 1–17 (2005).
9. Ishii, K. J. & Akira, S. Innate immune recognition of, and regulation by, DNA. *Trends Immunol.* **27**, 525–532 (2006).
10. Manji, G. A. *et al.* PYPAF1, a PYRIN-containing Apafl-like protein that assembles with ASC and regulates activation of NF- $\kappa$ B. *J. Biol. Chem.* **277**, 11570–11575 (2002).
11. Grenier, J. M. *et al.* Functional screening of five PYPAF family members identifies PYPAF5 as a novel regulator of NF- $\kappa$ B and caspase-1. *FEBS Lett.* **530**, 73–78 (2002).
12. Wang, L. *et al.* PYPAF7, a novel PYRIN-containing Apafl-like protein that regulates activation of NF- $\kappa$ B and caspase-1-dependent cytokine processing. *J. Biol. Chem.* **277**, 29874–29880 (2002).
13. Finn, R., Griffiths-Jones, S. & Bateman, A. Identifying protein domains with the Pfam database. *Curr. Protoc. Bioinformatics* **Chapter 2**, Unit-2.5 (2003).
14. Albrecht, M., Choubey, D. & Lengauer, T. The HIN domain of IFI-200 proteins consists of two OB folds. *Biochem. Biophys. Res. Commun.* **327**, 679–687 (2005).
15. Landolfo, S., Gariglio, M., Gribaudo, G. & Lembo, D. The Ifi 200 genes: an emerging family of IFN-inducible genes. *Biochimie* **80**, 721–728 (1998).
16. Ding, Y. *et al.* Antitumor activity of IFIX, a novel interferon-inducible HIN-200 gene, in breast cancer. *Oncogene* **23**, 4556–4566 (2004).
17. Trapani, J. A. *et al.* A novel gene constitutively expressed in human lymphoid cells is inducible with interferon- $\gamma$  in myeloid cells. *Immunogenetics* **36**, 369–376 (1992).
18. Burrus, G. R., Briggs, J. A. & Briggs, R. C. Characterization of the human myeloid cell nuclear differentiation antigen: relationship to interferon-inducible proteins. *J. Cell. Biochem.* **48**, 190–202 (1992).
19. DeYoung, K. L. *et al.* Cloning a novel member of the human interferon-inducible gene family associated with control of tumorigenicity in a model of human melanoma. *Oncogene* **15**, 453–457 (1997).
20. Yu, J. W. *et al.* Cryopyrin and pyrin activate caspase-1, but not NF- $\kappa$ B, via ASC oligomerization. *Cell Death Differ.* **13**, 236–249 (2006).
21. Szentesi, G. *et al.* Computer program for determining fluorescence resonance energy transfer efficiency from flow cytometric data on a cell-by-cell basis. *Comput. Methods Programs Biomed.* **75**, 201–211 (2004).
22. Henry, T. *et al.* Type I interferon signaling is required for activation of the inflammasome during *Francisella* infection. *J. Exp. Med.* **204**, 987–994 (2007).
23. Sun, K. H., Yu, C. L., Tang, S. J. & Sun, G. H. Monoclonal anti-double-stranded DNA autoantibody stimulates the expression and release of IL-1 $\beta$ , IL-6, IL-8, IL-10 and TNF- $\alpha$  from normal human mononuclear cells involving in the lupus pathogenesis. *Immunology* **99**, 352–360 (2000).
24. Schachman, H. K. *et al.* Enzymatic synthesis of deoxyribonucleic acid. VII. Synthesis of a polymer of deoxyadenylate and deoxythymidylate. *J. Biol. Chem.* **235**, 3242–3249 (1960).
25. Kanneganti, T. D. *et al.* Critical role for Cryopyrin/Nalp3 in activation of caspase-1 in response to viral infection and double-stranded RNA. *J. Biol. Chem.* **281**, 36560–36568 (2006).
26. Severa, M., Coccia, E. M. & Fitzgerald, K. A. Toll-like receptor-dependent and -independent viperin gene expression and counter-regulation by PRDI-binding factor-1/BLIMP1. *J. Biol. Chem.* **281**, 26188–26195 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We would like to thank A. Cerny for animal husbandry and genotyping and R. Johnstone for the anti-AIM2 antibody. V.H. is supported by a fellowship from the Deutsche Forschungsgemeinschaft (German Research Foundation; Ho2783/2-1), E.L. and K.A.F. are supported by grants from the National Institutes of Health (AI-065483 (to E.L.) and AI-067497 (to K.A.F.)).

**Author Contributions** V.H. conceived the research and conducted the experiments with A.A., M.C.-D., F.B. and G.H. D.R.C. performed sequence analysis. E.L. and K.A.F. oversaw the whole project.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to K.A.F. ([kate.fitzgerald@umassmed.edu](mailto:kate.fitzgerald@umassmed.edu)) or V.H. ([veit.hornung@uni-bonn.de](mailto:veit.hornung@uni-bonn.de)).

## METHODS

**Plasmid constructs.** Full-length human AIM2 (1–356, on the basis of BC010940.1), AIM2-PYD (1–83), AIM2-HIN (148–356), IFIX (1–461, NM\_198929), IFIX-PYD (1–84), IFI16 (1–729, NM\_005531), IFI16-PYD (1–84), MND A (1–407), MND A-PYD (1–84, NM\_002432) and NLRP3-PYD (1–87, NM\_004895.3) were cloned by PCR from cDNA into pEFBOS-C-term-CFP using PCR-generated XhoI and BamHI or BglII restriction sites. AIM2 full-length, AIM2-PYD and AIM2-HIN were subcloned into pEFBOS-C-term-Flag/His using XhoI and BamHI. Murine pro-IL1 $\beta$  (1–269) was obtained by PCR from cDNA and fused into pEFBOS-C-term-GLuc/Flag using XhoI and BglII/BamHI. Expression plasmids (pCI) encoding human caspase-1 and ASC-HA were from Millenium Pharmaceuticals.

**Reagents.** ATP, LPS and poly(dA:dT) were from Sigma-Aldrich. A555-conjugated cholera-toxin B was from Molecular Probes, Invitrogen. DRAQ5 was from Biostatus. Biotinylated and FITC-labelled poly(dA:dT) were made by adding biotin-dUTP or FITC-dUTP (Fermentas) at a molar ratio of 1:8 to dTTP in the enzymatic synthesis of poly(dA:dT) as described<sup>24</sup>. Anthrax lethal toxin (anthrax protective antigen and lethal factor) was from List Biologicals). The anti-human-AIM2 antibody (3B10) was from R. Johnstone.

**Mice.** *Nlrp3*<sup>−/−</sup> and *Pycard*<sup>−/−</sup> mice were as previously described<sup>25</sup>. Both strains, as well as *Nlrp6*<sup>−/−</sup> and *Nlrp12*<sup>−/−</sup> mice were from Millennium Pharmaceuticals, and caspase-1-deficient mice were from R. Flavell. C57BL/6, 129/Sv, C57/BL6  $\times$  129 F<sub>1</sub> and *Il1r1*<sup>−/−</sup> mice were from Jackson Laboratories. All mouse strains were bred and maintained under specific pathogen-free conditions in the animal facilities at the University of Massachusetts Medical School.

**Sequence analysis.** Pyrin-domain-containing sequences were retrieved from UniProt after identifying them in SMART. A multiple-sequence alignment of selected pyrin domains was generated using MUSCLE<sup>27</sup>. Secondary structure elements from the POP1 crystal structure (Protein Data Bank accession 2HM2) were mapped to the multiple-sequence alignment in PFAAT<sup>28</sup>. Prediction of nuclear targeting sequences and subcellular localization was done using PSORTII (<http://psort.ims.u-tokyo.ac.jp/>).

**Cell culture and stimulation.** Bone-marrow-derived macrophages were generated and cultured in DMEM medium as described<sup>26</sup>. THP-1 cells and macrophage cell lines were cultured as previously described<sup>7</sup>. One day before stimulation, THP-1 cells were differentiated using 0.5  $\mu$ M PMA<sup>7</sup>. ATP (5 mM) was added 1 h before collection of supernatants. Poly(dA:dT) DNA and all other DNAs were transfected using Lipofectamine 2000 at a concentration of 1  $\mu$ g ml<sup>−1</sup> according to the manufacturer's instructions. Expression plasmids were transfected into 293T cells using GeneJuice (Novagen). Vaccinia virus (WR strain) was used for infection at a MOI of 5 if not indicated otherwise. Anthrax lethal toxin (protective antigen and lethal factor) were both at 5  $\mu$ g ml<sup>−1</sup>.

**ELISA.** Cell culture supernatants were assayed for IL1 $\beta$  by ELISA (BD Biosciences). To measure intracellular IL1 $\beta$ , cells were washed and subjected to three freeze-thaw cycles in assay diluent.

**Confocal microscopy.** Confocal microscopy was performed on a Leica SP2 AOBs confocal laser scanning microscope. Separation of CFP and YFP was performed using sequential scanning and 405 and 514 nm excitation.

**Flow cytometry fluorescence resonance energy transfer.** Three fluorescent intensities were measured: I1, direct excitation and emission of donor; I2, indirect excitation and direct emission of acceptor; and I3, direct excitation and emission of acceptor. After full correction for spectral bleed-through and cross excitation, FRET efficiency was calculated on a cell-by-cell basis<sup>21</sup> and then FRET efficiency histograms were plotted with GraphPad Prism 5.01 (GraphPad Software).

**Immunoblot analysis.** Immunoblotting was conducted as previously described<sup>7</sup> using anti-murine caspase-1 p10 (Santa Cruz Biotechnology), anti-Flag (M2, Sigma), anti-HA (Roche Applied Science), anti-CFP (Santa Cruz Biotechnology), anti-ASC (Alexis, AL177), anti-IFI16 (sc-8023, Santa Cruz Biotechnology) or anti-AIM2 (3B10 mouse IgG1<sup>29</sup>).

**Co-immunoprecipitation assays.** 293T cells (24 wells) were transfected with 1,550 ng of the PYD-CFP expression plasmids and 50 ng of ASC-HA. Twenty-four hours later, cell lysates were cleared by centrifugation (20,000g, 30 min) and subsequently incubated with anti-HA agarose beads for 2 h at 4 °C. Differentiated THP-1 cells were primed with Sendai virus (300 haemagglutinating units (HAU) ml<sup>−1</sup>) overnight and subsequently lysed in high salt lysis buffer (250 mM NaCl, 10 mM Tris-HCl, pH 7.4, 1% CHAPS, protease inhibitor cocktail). Lysates were cleared by centrifugation, salt concentration was adjusted to 125 mM NaCl and ASC was immunoprecipitated using rabbit polyclonal anti-ASC antibody. After six washes in both cases beads were boiled with Laemmli buffer for immunoblot analysis.

**Quantitative real-time PCR.** Quantitative rtPCR analysis was performed as described<sup>26</sup>. Primer sequences for murine *Hprt1*, *Ifnb*, *Aim2* and *Il6* are available upon request.

**shRNA-mediated silencing.** The lentiviral shRNA expression plasmids were from OpenBiosystems. The shRNAs targeting *Aim2* are: TRCN0000096104 (shRNA1), TRCN0000096105 (shRNA2) and TRCN0000096106 (shRNA3). The control shRNA is directed against murine *Ifih1* (TRCN0000103648) and was confirmed not to have any effect on *Nlrp3* or *Aim2* expression. The production of viral particles and transduction of target cells was conducted as described on [http://www.broad.mit.edu/genome\\_bio/trc/publicProtocols.html](http://www.broad.mit.edu/genome_bio/trc/publicProtocols.html).

**AlphaScreen assay.** The AlphaScreen was set up as an association assay. Proteins were transiently expressed in 293T cells and purified using Flag beads (Sigma) binding to the C-terminal Flag/His tag. The protein of interest was incubated at a concentration of 100 nM with biotinylated ligand at the indicated concentration in PBS, 0.1% BSA and 0.01% Tween 20 for 60 min. Subsequently, nickel chelate acceptor beads (binding to the His tag) and streptavidin-coated donor beads were added. After 30 min incubation at 25 °C in the dark, samples were read in proximity plates with the Envision HT microplate reader (all Perkin Elmer). Data were analysed by GraphPad Prism.

**Reporter assays.** All reporter gene assays were conducted as described<sup>26</sup>.

**siRNA transfection.** Differentiated THP-1 cells were plated at  $2.5 \times 10^4$  cells per well in 96-well plates and transfected with siRNA targeting human *AIM2* (sense strand: 5'-CCCGAAGATCAACACGCTTCA-3'), human *ASC* (sense strand: 5'-CGGGAAGGTCTGACGGATGA-3') or human *TLR8* (5'-GGGAGUUA-CUGCUUGAAG A-3') using 275 ng siRNA and 0.75  $\mu$ l Lipofectamine 2000. Cells were stimulated as indicated 48 h after transfection.

**Cell viability assay.** To quantify cell viability, macrophages ( $1 \times 10^5$  cells per well in 96-well plates) were treated as described. After 24 h, cells were washed with PBS and incubated in PBS plus 5  $\mu$ M calcein AM (Invitrogen) for 30 min at 37 °C. The number of viable cells was assessed by counting fluorescent cells in two independent visual fields ( $\times 20$  magnification) using ImageJ or by determining the overall fluorescence intensity using an Envision HT microplate reader.

27. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

28. Caffrey, D. R. *et al.* PFAAT version 2.0: a tool for editing, annotating, and analyzing multiple sequence alignments. *BMC Bioinformatics* **8**, 381 (2007).

29. Cresswell, K. S. *et al.* Biochemical and growth regulatory activities of the HIN-200 family member and putative tumor suppressor protein, AIM2. *Biochem. Biophys. Res. Commun.* **326**, 417–424 (2005).

# Dynamic expression of epidermal caspase 8 simulates a wound healing response

Pedro Lee<sup>1</sup>, Dai-Jen Lee<sup>1</sup>, Carol Chan<sup>1</sup>, Shih-Wei Chen<sup>1</sup>, Irene Ch'en<sup>2</sup> & Colin Jamora<sup>1,3</sup>

Tissue homeostasis and regeneration are regulated by an intricate balance of seemingly competing processes—proliferation versus differentiation, and cell death versus survival<sup>1</sup>. Here we demonstrate that the loss of epidermal caspase 8, an important mediator of apoptosis<sup>2</sup>, recapitulates several phases of a wound healing response in the mouse. The epidermal hyperplasia in the caspase 8 null skin is the culmination of signals exchanged between epidermal keratinocytes, dermal fibroblasts and leukocytic cells. This reciprocal interaction is initiated by the paracrine signalling of interleukin 1 $\alpha$  (IL1 $\alpha$ ), which activates both skin stem cell proliferation and cutaneous inflammation. The non-canonical secretion of IL1 $\alpha$  is induced by a p38-MAPK-mediated upregulation of NALP3 (also known as NLRP3), leading to inflammasome assembly and caspase 1 activation. Notably, the increased proliferation of basal keratinocytes is counterbalanced by the growth arrest of suprabasal keratinocytes in the stratified epidermis by IL1 $\alpha$ -dependent NF $\kappa$ B signalling. Altogether, our findings illustrate how the loss of caspase 8 can affect more than programmed cell death to alter the local microenvironment and elicit processes common to wound repair and many neoplastic skin disorders.

The structure and function of the epidermis is maintained by an equilibrium between proliferating keratinocytes in the basal layer and differentiation of cells into the suprabasal layers that seal the body from the environment (Supplementary Fig. 1a). A further regulatory balance occurs at the epidermal surface between keratinocyte survival and death, which requires the continual regeneration of this tissue. During wound healing these respective equilibria shift towards keratinocyte proliferation and cell survival to stimulate repair of the damaged tissue. Despite their importance, the mechanisms governing these regulatory nodes of tissue homeostasis remain largely undefined.

In the course of characterizing proteins that potentially affect epidermal homeostasis, we observed that epidermal caspase 8 normally fluctuates during the course of a wound healing response (Fig. 1a). *In situ* hybridization revealed that sites proximal (<1 mm) to an excisional wound showed a thickened epidermis coincident with a down-regulation of caspase 8 RNA. However at sites distal (~5 mm) to the wound, in which epidermal thickness appeared normal, caspase 8 RNA was unchanged. After wound closure (day 14) epidermal hyperplasia subsides and caspase 8 RNA expression was restored. The development of all epidermal layers before the onset of caspase 8 expression supports the notion that this protein is not involved in epidermal morphogenesis but instead monitors epidermal integrity amidst assaults from the external environment (Supplementary Fig. 1b).

To define the contribution of reduced caspase 8 to a wound healing response, we generated an epidermal knockout of caspase 8 (refs 3, 4). By postnatal day 10 (P10) the conditional knockout mouse has flaky skin throughout its body, is slightly runted and its epidermis is dramatically

thickened (Supplementary Fig. 2a, b). We unexpectedly found that even though caspase 8 is expressed in the granular layer (Supplementary Fig. 1c), it is the basal and spinous layers that are markedly expanded in the knockout epidermis (Fig. 1b). Keratinocyte proliferation contributes to this phenotype as Ki67 (Fig. 1c) and cyclin D (Supplementary Fig. 3a) are increased along the basement membrane in the knockout skin. This observation is compatible with data showing that both epidermal and hair follicle stem cells are mobilized to help reseal a wounded epidermis<sup>5–7</sup>.

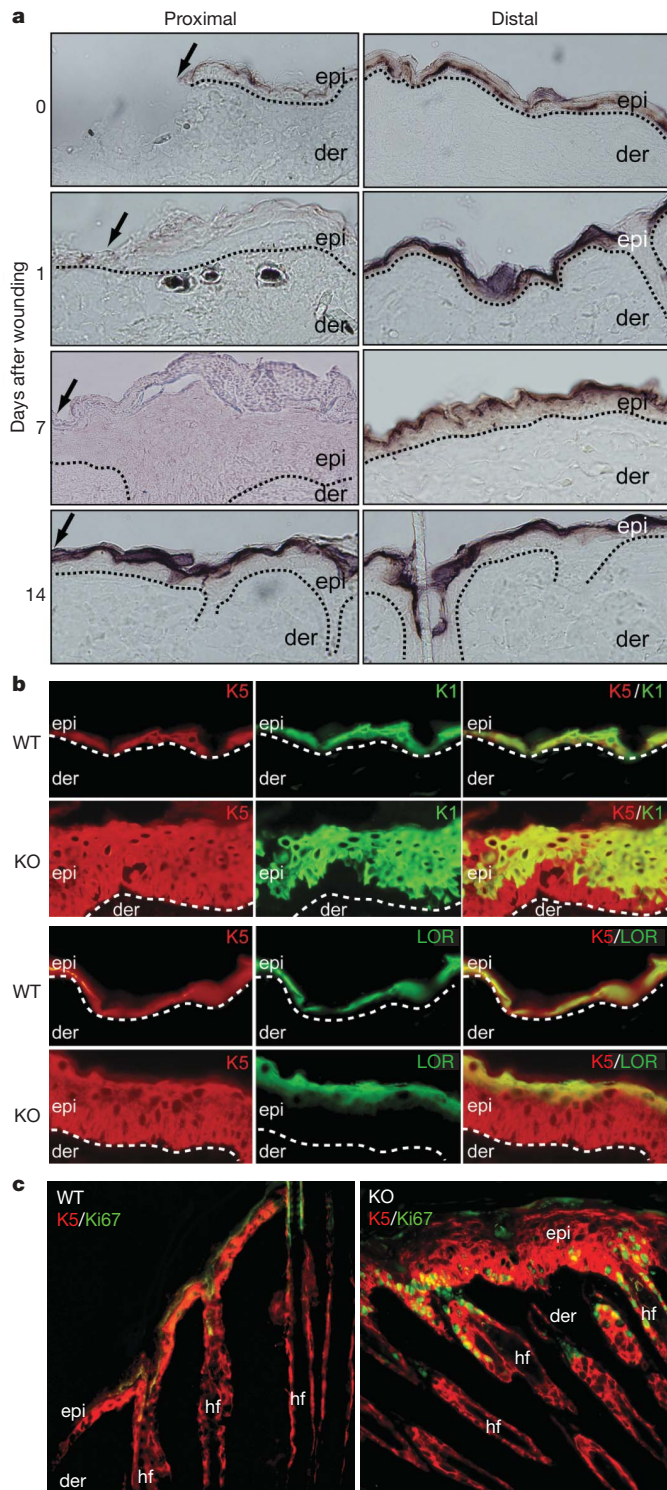
Other markers of epidermal hyperproliferation that occur during wound healing are upregulated in the caspase 8 knockout skin including keratin 6 (K6, also known as KRT6) and suprabasal expression of  $\beta$ 4 integrin (Supplementary Fig. 3b, c). Despite this shift in balance from differentiation to proliferation, the terminal differentiation program nevertheless progressed to completion (Supplementary Fig. 3d). Together these observations suggest that the loss of caspase 8 near the epidermal surface elicits a paracrine signal(s) to stimulate local stem cell pools to proliferate and fuel wound closure.

Among the early phases of a wound healing response is the infiltration of immune cells at the wound site<sup>8</sup>. The increased dermal cellularity in the knockout skin (Supplementary Fig. 2b) suggests that this is also occurring in response to the loss of caspase 8. Analysis with different leukocyte markers demonstrates that the caspase 8 knockout skin contains a marked increase in the population of granulocytes and macrophages (Fig. 2a). Probing with the pan T-cell marker CD3 (also known as CD3E) revealed increased levels of T cells in both the epidermis and dermis (Fig. 2b). Both CD4 helper T cells and CD8 cytotoxic T cells were increased in the epidermal and dermal compartments (Fig. 2b). Owing to the pivotal part they play in wound healing, we investigated the status of  $\gamma\delta$  T cells<sup>9</sup>. In wild-type skin, these cells are restricted to the basal layer and hair follicle and have a dendritic morphology (Fig. 2b). In the caspase 8 knockout skin, the number of  $\gamma\delta$  T-cells is increased in the suprabasal layers and dermis, with a spherical morphology indicative of their activation. The  $\gamma\delta$  T cells in the dermis stain positive for V $\gamma$ 3 (Fig. 2b), suggesting that they originated from cells that proliferated and migrated from the epidermis<sup>9</sup>. These immune cells were mediating an inflammatory reaction similar to a wound healing response as evidenced by the induction of NF $\kappa$ B-stimulated inflammatory genes (Fig. 2c). Unlike other mutant mice with hyperplasia and inflammation phenotypes<sup>10,11</sup>, suppression of the immune cell infiltration did not completely abrogate cell proliferation (Supplementary Fig. 4a–d). The diminished hyperplasia in the presence of immunosuppressants suggests that epidermal and immune cell signalling work together to maximally stimulate keratinocyte proliferation.

From these observations, the question arises as to how the loss of caspase 8 near the epidermal surface only stimulates proliferation in the cells contiguous with the basal layer. A clue was provided by the

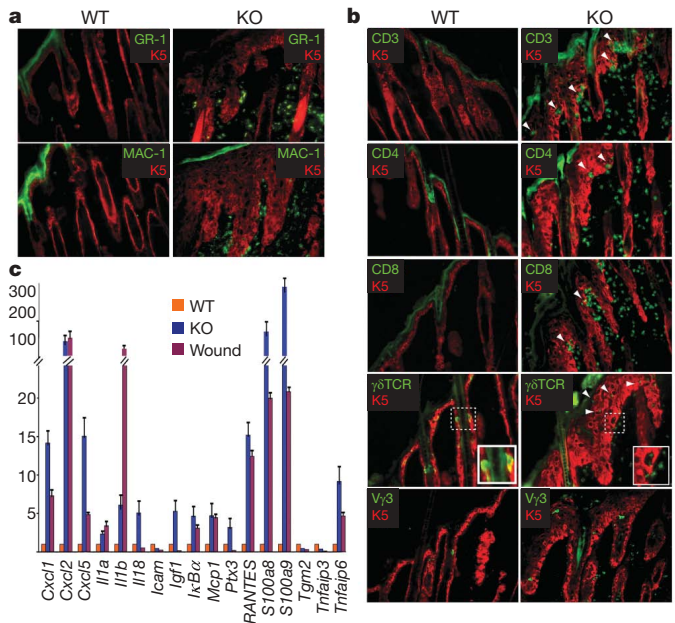
<sup>1</sup>Section of Cell and Developmental Biology, Division of Biological Sciences, Natural Science Building, Room 6311, 9500 Gilman Drive, MC 0380, <sup>2</sup>Section of Molecular Biology, Division of Biological Sciences, Natural Science Building, Room 5116, 9500 Gilman Drive, MC 0377, <sup>3</sup>Department of Medicine (Dermatology), Natural Science Building, Room 6113, 9500 Gilman Drive, MC 0380, La Jolla, California 92093, USA.





**Figure 1 | Effect of caspase 8 downregulation in the epidermis.** **a**, *In situ* hybridization of caspase 8 RNA on days 0, 1, 7 and 14 after excisional wounds. Arrows denote wound sites. Distal (>5 mm) and proximal (<1 mm) refer to the distance from the wound site. der, dermis; epi, epidermis. **b**, Differentiation in the wild type (WT) and caspase 8 knockout (KO) epidermis. K5, K1 (also known as KRT5 and KRT1, respectively) and loricrin (LOR) stain the basal, spinous and granular layers, respectively. Dotted lines denote the basement membrane. **c**, Hyperproliferation of the epidermis in the knockout skin is revealed by increased expression of Ki67. hf, hair follicle. The original magnification used for all images was  $\times 40$ .

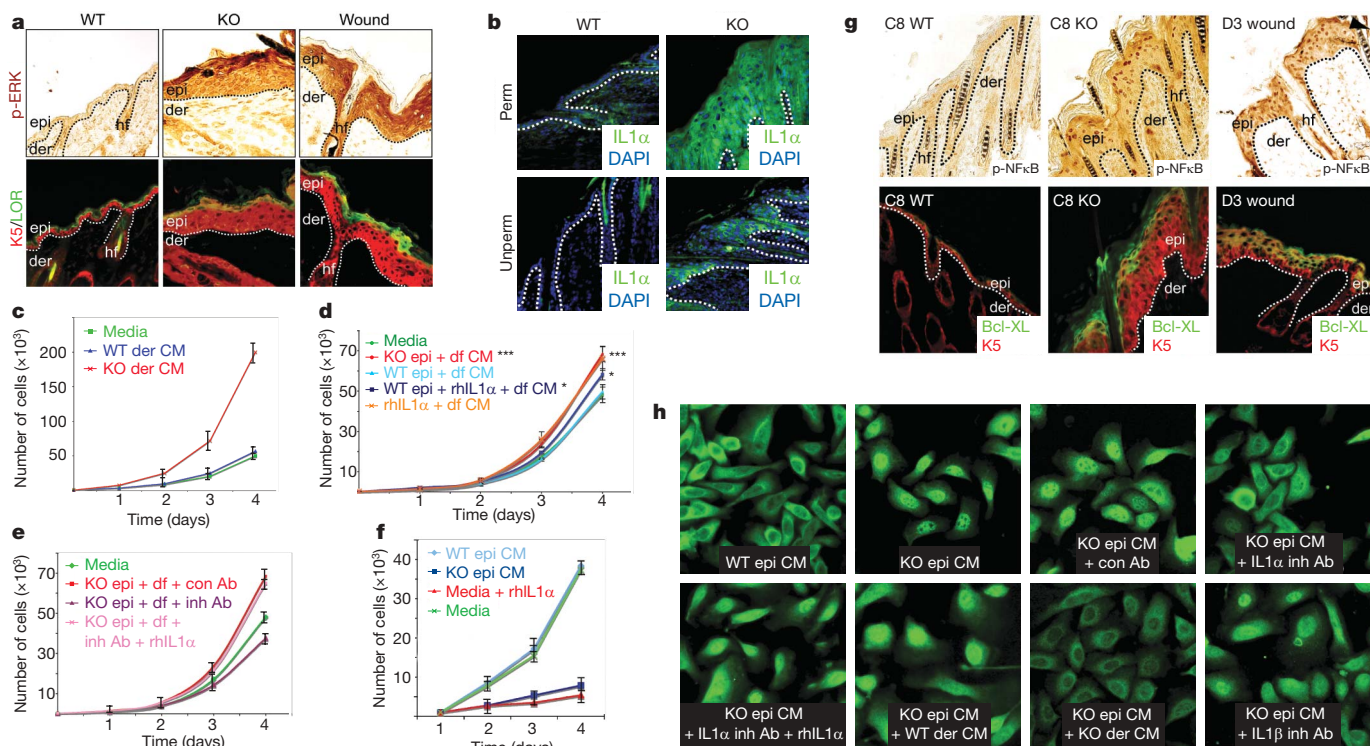
observation that phosphorylated ERK (p-ERK; Fig. 3a and Supplementary Fig. 5a) and  $\beta 4$  integrin (Supplementary Fig. 3c) were increased in the knockout tissue and expressed in the suprabasal layers.



**Figure 2 | Characterization of the inflammatory response in the caspase 8 knockout skin.** **a**, Increased number of granulocytes (GR-1, also known as LY6G) and macrophages (MAC-1, also known as ITGB2) are detected in knockout (KO) skin. **b**, Pan (CD3), helper (CD4), cytotoxic (CD8) and  $\gamma\delta$  ( $\gamma\delta$ TCR, recognizing the  $\gamma\delta$  T-cell receptor) T-cell markers are increased in the knockout skin compared to wild type (WT) and infiltrate the epidermis (arrowheads). Insets show magnified views of  $\gamma\delta$  T-cell morphology. The epidermal-derived pool of  $\gamma\delta$  T cells is stained with V $\gamma$ 3. **c**, Quantitative PCR with reverse transcriptase (qRT-PCR) of pro-inflammatory NF $\kappa$ B target genes in wild type (orange) and knockout (blue) skin of 10-day-old mice and wounded skin (purple) samples 3 days after wounding. The original magnification used for all images was  $\times 40$ . Error bars are s.e.m.

This is noteworthy because transgenic mice designed to suprabasally express active  $\beta 4$  integrin or MEK1 (also known as MAP2K1; the upstream activator of ERK) phenocopy the caspase 8 knockout mouse<sup>12,13</sup>. Both mouse models have an increased level of active IL1 which has an established role in regulating keratinocyte proliferation and triggering cutaneous inflammation<sup>14,15</sup>. Because IL1 $\beta$  is primarily secreted from leukocytes, we focused our investigation on the activity of IL1 $\alpha$  which is highly expressed in keratinocytes<sup>16</sup> and is secreted from the knockout epidermis (Fig. 3b). It has been proposed that epidermally derived IL1 $\alpha$  mediates keratinocyte proliferation by initially stimulating fibroblasts of the underlying dermis to secrete cytokines that subsequently incite keratinocyte division<sup>15</sup>. Consistent with this hypothesis, using normal serum levels we found that conditioned media from the dermis of knockout animals stimulated keratinocytes to proliferate at a faster rate than those exposed to wild-type conditioned or normal culture media (Fig. 3c). Because the dermal explants are a heterogeneous population of cells, we investigated whether dermal fibroblasts are singularly competent to stimulate keratinocyte proliferation. Given our hypothesis that IL1 $\alpha$  is provided by the epidermis, we first primed the dermal fibroblasts with media exposed to epidermal sheets and then treated keratinocytes with this conditioned media. Priming the dermal fibroblasts with knockout epidermal media caused a 40% increase in keratinocyte growth compared to the effects of either wild-type epidermal-primed dermal fibroblast media or regular keratinocyte culture media (Fig. 3d). Adding recombinant human IL1 $\alpha$  (rhIL1 $\alpha$ ) directly to the dermal fibroblast was sufficient to induce them to secrete proliferation-stimulating cytokines, whereas an IL1 $\alpha$ -neutralizing antibody abolished the ability of the knockout epidermis to prime the dermal fibroblasts (Fig. 3d, e). These data suggest that IL1 $\alpha$  is a required component secreted by the epidermal sheet to render the dermal fibroblast capable of stimulating keratinocyte proliferation.

Notably, we found that either epidermal conditioned media alone or rhIL1 $\alpha$  can stunt keratinocyte growth (Fig. 3f). We therefore



**Figure 3 | Control of keratinocyte proliferation through**

**epithelial-mesenchymal interactions.** **a**, Expression of p-ERK, LOR and K5 in serial sections from postnatal day (P10) wild-type (WT), knockout (KO) and skin 3 days after wounding. **b**, Extracellular (unperm, unpermeabilized) and total (perm, permeabilized) expression levels of IL1 $\alpha$ . **c**, Keratinocyte growth after incubation in conditioned media from wild-type dermal (WT der CM; blue), caspase 8 knockout dermal (KO der CM; red) or normal growth media (green). **d**, Keratinocyte growth after incubation in normal media (green), conditioned media from dermal fibroblasts primed with wild-type epidermis-exposed media (WT epi + df CM; light blue) or primed with caspase 8 null epidermal media (KO epi + df CM; red), dermal fibroblast plus rhIL1 $\alpha$  (rhIL1 $\alpha$  + df CM; orange), and conditioned media from dermal fibroblasts primed with wild-type epidermal-media containing rhIL1 $\alpha$  (WT epi + rhIL1 $\alpha$  + df CM; dark blue). \* $P < 0.05$  and \*\*\* $P < 0.001$  compared to normal media. **e**, Keratinocyte growth after incubation in normal media (green), dermal fibroblast media primed with knockout epidermal media and control antibody (KO epi + df + con Ab; red), IL1 $\alpha$

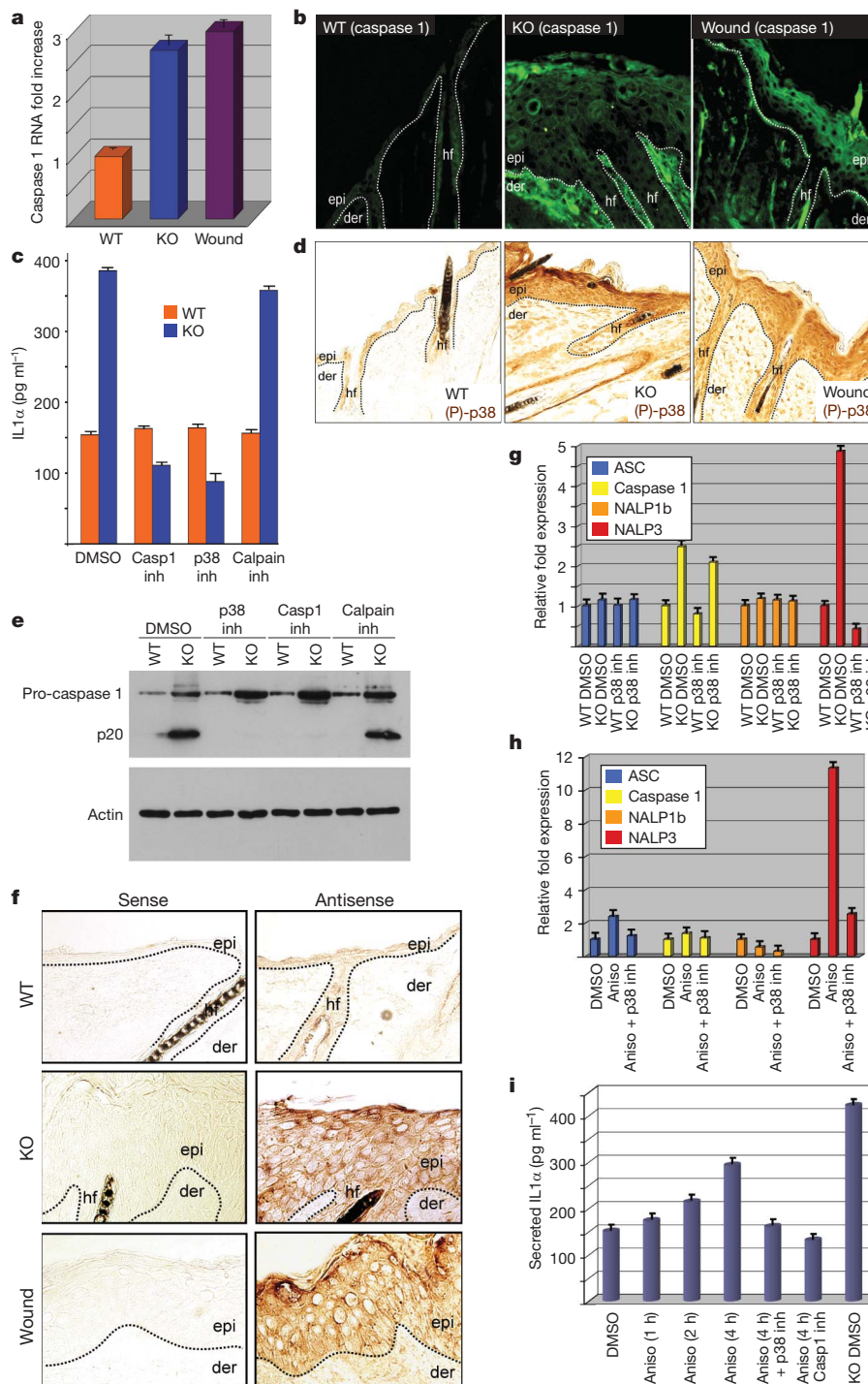
inhibitory antibody (KO epi + df + inh Ab; purple), or rhIL1 $\alpha$  (KO epi + df + inh Ab + rhIL1 $\alpha$ ; pink). **f**, Wild-type epidermal-conditioned media (light blue), knockout epidermal conditioned media (dark blue) or normal media (green) was added to primary keratinocytes and growth rate was measured. Addition of rhIL1 $\alpha$  to normal media (red) is sufficient to inhibit growth. **g**, Skin sections from P10 caspase 8 (C8) wild-type and knockout mice and adult mouse skin 3 days (D3) after excisional wounding were stained for p-NF $\kappa$ B, its target gene Bcl-XL (green) and K5 (red). Arrowhead marks wound site. **h**, Wild-type and knockout epidermal conditioned media were added to primary keratinocytes and stained for NF $\kappa$ B nuclear translocation. Inhibitory IL1 $\alpha$  antibody, which specifically neutralizes the mouse cytokine, and/or recombinant human IL1 $\alpha$  (rhIL1 $\alpha$ ) was added to the knockout media. Keratinocytes were also treated with knockout epidermal media, wild-type dermal media, knockout dermal media or a neutralizing antibody specific for mouse IL1 $\beta$ . Results are representative of three independent experiments. The original magnification used for all images was  $\times 40$ . All error bars are s.d.

investigated the possibility that IL1 $\alpha$  may induce NF $\kappa$ B signalling, which is known to promote growth inhibition and cell survival in keratinocytes<sup>17,18</sup> and found increased nuclear phosphorylated-NF $\kappa$ B (p-NF $\kappa$ B) in the knockout and wounded epidermis (Fig. 3g). Bcl-XL (also known as BCL2L1; Fig. 3g) as well as other NF $\kappa$ B transcriptional target genes (Supplementary Fig. 6) were similarly increased. To determine whether cell growth inhibition was due to IL1 $\alpha$  signalling, we stained for nuclear translocation of NF $\kappa$ B in response to cells treated with conditioned media from wild-type and knockout epidermal sheets (Fig. 3h). Keratinocytes treated with wild-type conditioned media had NF $\kappa$ B predominantly localized in the cytoplasm, whereas cells treated with knockout media showed nuclear localization. The nuclear translocation could be inhibited when knockout conditioned media was added together with an IL1 $\alpha$ -neutralizing antibody and rescued with rhIL1 $\alpha$  (which is not recognized by the inhibitory antibody; Fig. 3h). Moreover, the addition of knockout dermis conditioned media can overcome the growth inhibitory effects of the knockout epidermal conditioned media and block the nuclear translocation of NF $\kappa$ B, which is consistent with the former's ability to stimulate keratinocyte proliferation. Thus the induction of NF $\kappa$ B in the suprabasal layers is a contributing factor towards limiting proliferation to cells along the basement membrane in the knockout epidermis.

Given the complex effects of IL1 $\alpha$  signalling in the mutant mouse, it became imperative to determine how the loss of caspase 8 regulates the secretion (and therefore the activity) of this cytokine. Two proteases have been implicated in the conversion of the interleukin 1 precursor into its active form: caspase 1 and calpain<sup>18</sup>. In both the caspase 8 null epidermis and wounded skin there is an upregulation of caspase 1 RNA and suprabasal localization of the protein (Fig. 4a, b). Furthermore, the amount of IL1 $\alpha$  secreted from epidermal sheets is 2.5-fold higher in the knockout mouse than its wild-type littermate and this increase is lost when treated with a caspase 1 (but not the calpain) inhibitor (Fig. 4c). These results are supported by reports that the overexpression of caspase 1 (ref. 19) or IL1 $\alpha$  (ref. 20) in the epidermis phenocopies the caspase 8 conditional knockout.

Caspase 1 is also synthesized as a zymogen and must be processed into its active form. We found that phosphorylated p38 MAPK is upregulated in both the caspase 8 knockout and wounded skin (Fig. 4d and Supplementary Fig. 5b) and inhibition of its activity abolishes the secretion of IL1 $\alpha$  from the knockout epidermis (Fig. 4c). The p38 MAPK inhibitor did not abrogate the increase in caspase 1 protein expression in the knockout epidermis, but did efficiently restrain caspase 1 processing into its active form and thus its catalytic activity (Fig. 4e and Supplementary Fig. 8a). A consequence of blocking p38-MAPK-mediated activation of caspase 1





**Figure 4 | Regulation of IL1 $\alpha$  secretion.** **a**, qRT-PCR for caspase 1 in P10 wild-type (WT), knockout (KO) and wounded skin 3 days after excision. **b**, Immunofluorescence expression of caspase 1 in P10 wild type, knockout and wounded skin. **c**, IL1 $\alpha$  ELISA from media conditioned by wild type (orange) or knockout (blue) epidermal sheets treated with either dimethylsulphoxide (DMSO), or inhibitors (inh) for caspase 1 (casp1), p38 MAPK or calpain. The calpain inhibitor efficacy was verified by its ability to inhibit processing of filaggrin in differentiating keratinocytes (Supplementary Fig. 7). **d**, Expression of phosphorylated p38 MAPK ((P)-p38) in P10 wild-type and knockout skin and 3 days after wounding. **e**, Western blot of caspase 1 and  $\beta$ -actin from cell lysates generated from the epidermal sheets in **c**. **f**, *In*

*situ* hybridization of NALP3 RNA on P10 wild-type, knockout and 3 days post-wounded skin sections. **g**, qRT-PCR of inflammasome components using RNA isolated from epidermal sheets of wild-type and knockout skin treated with DMSO vehicle control, anisomycin (aniso), or p38 MAPK inhibitor for 16 h. **h**, qRT-PCR on RNA from wild-type epidermal sheets treated with DMSO vehicle control, anisomycin, or anisomycin plus p38 MAPK inhibitor for 2 h. **i**, Measurement of IL1 $\alpha$  secretion in conditioned media from wild-type epidermal sheets treated with DMSO, anisomycin, or anisomycin plus p38 MAPK inhibitor or caspase 1 inhibitor. Knockout epidermal sheets treated with DMSO are used as a comparison. The original magnification used for all images was  $\times 40$ . All error bars are s.e.m.

is the reduction in epidermally derived IL1 $\alpha$  signalling as measured by nuclear translocation of NF $\kappa$ B (Supplementary Fig. 8b).

To probe the mechanistic link between p38 MAPK and caspase 1, we analysed the status of the NALP3 inflammasome, a multiprotein

complex which is required for activation of inflammatory caspases such as caspase 1 (ref. 21). This complex is composed of the adaptor protein ASC (also known as PYCARD) and pro-caspase 1—both of which have a basal expression level in keratinocytes—and NALP3, the



expression of which must be induced<sup>21</sup>. We found that *NALP3* RNA is markedly upregulated in the suprabasal layers of both the knockout and wounded epidermis (Fig. 4f). This expression pattern coincides with active p38 MAPK expression in the knockout epidermis and we found that *NALP3* induction is dependent on the activity of this kinase (Fig. 4g). Furthermore, treatment of epidermal sheets with the p38 MAPK chemical activator anisomycin is sufficient to induce *NALP3* expression in wild-type epidermal sheets (Fig. 4h). The human NLRP1 (previously known as *NALP1*) protein can also form an inflammasome complex with caspase 1 to mediate its proteolytic activation, but its closest murine relative, *NALP1b* (*NLRP1B*), is unaffected in the caspase 8 knockout epidermis and is unaltered in response to modulating p38 MAPK activity (Fig. 4g, h). Consistent with its role in stimulating inflammasome-dependent activation of caspase 1, anisomycin can increase the amount of IL1 $\alpha$  secreted from epidermal sheets (Fig. 4i). The ability of both ultraviolet-irradiated<sup>22</sup> and anisomycin-treated<sup>23</sup> keratinocytes to expel their inflammasome components prompted us to determine whether p38 MAPK can also affect the secretion of caspase 1. Conditioned media from caspase 8 null epidermal sheets contained increased levels of released pro-caspase 1, which was reduced to wild-type levels when p38 MAPK activity was compromised (Supplementary Fig. 8c). The p20 activated fragment of caspase 1 was also higher in the extracellular media from the knockout epidermis, and both p38 MAPK and caspase 1 inhibition abolished the release of this protein from the cell.

A model summarizing how a wound-induced downregulation of epidermal caspase 8 is translated into epithelial–mesenchymal–leukocyte interactions is shown in Supplementary Fig. 9. Consistent with this proposed role, caspase 8 is increased in diabetic mice with impaired wound healing<sup>24</sup> and humans deficient in caspase 8 have clinical symptoms such as eczema, which has features of a chronic wound response<sup>25</sup>. In addition to its effect on wound healing, the intercellular communication stimulated by the loss of epidermal caspase 8 and the activation of the inflammatory apparatus may be a mechanism for tumour–stroma interactions and other hyperplastic skin diseases.

## METHODS SUMMARY

**Generation of caspase 8 conditional knockout mice.** Epidermal-specific knockouts were obtained by crossing mice carrying the floxed caspase 8 allele (*Casp8<sup>fl/j</sup>*)<sup>4</sup> to K14-Cre mice<sup>3</sup>.

**Histology, *in situ* hybridization and immunohistochemistry.** Please refer to the Methods.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 13 August; accepted 9 December 2008.

Published online 8 February 2009.

- Alonso, L. & Fuchs, E. Stem cells of the skin epithelium. *Proc. Natl Acad. Sci. USA* **100** (Suppl. 1), 11830–11835 (2003).
- Raj, D., Brash, D. E. & Grossman, D. Keratinocyte apoptosis in epidermal development and disease. *J. Invest. Dermatol.* **126**, 243–257 (2006).
- Vasioukhin, V., Degenstein, L., Wise, B. & Fuchs, E. The magical touch: genome targeting in epidermal stem cells induced by tamoxifen application to mouse skin. *Proc. Natl Acad. Sci. USA* **96**, 8551–8556 (1999).

- Beisner, D. R., Ch'en, I. L., Kolla, R. V., Hoffmann, A. & Hedrick, S. M. Cutting edge: innate immunity conferred by B cells is regulated by caspase-8. *J. Immunol.* **175**, 3469–3473 (2005).
- Ito, M. *et al.* Stem cells in the hair follicle bulge contribute to wound repair but not to homeostasis of the epidermis. *Nature Med.* **11**, 1351–1354 (2005).
- Levy, V., Lindon, C., Harfe, B. D. & Morgan, B. A. Distinct stem cell populations regenerate the follicle and interfollicular epidermis. *Dev. Cell* **9**, 855–861 (2005).
- Cotsarelis, G. Epithelial stem cells: a folliculocentric view. *J. Invest. Dermatol.* **126**, 1459–1468 (2006).
- Martin, P. & Leibovich, S. J. Inflammatory cells during wound repair: the good, the bad and the ugly. *Trends Cell Biol.* **15**, 599–607 (2005).
- Jameson, J. & Havran, W. L. Skin  $\gamma\delta$  T-cell functions in homeostasis and wound healing. *Immunol. Rev.* **215**, 114–122 (2007).
- Ichinohe, M. *et al.* Lack of phospholipase C- $\delta$ 1 induces skin inflammation. *Biochem. Biophys. Res. Commun.* **356**, 912–918 (2007).
- Perez-Moreno, M. *et al.* p120-catenin mediates inflammatory responses in the skin. *Cell* **124**, 631–644 (2006).
- Hobbs, R. M. & Watt, F. M. Regulation of interleukin-1 $\alpha$  expression by integrins and epidermal growth factor receptor in keratinocytes from a mouse model of inflammatory skin disease. *J. Biol. Chem.* **278**, 19798–19807 (2003).
- Hobbs, R. M., Silva-Vargas, V., Groves, R. & Watt, F. M. Expression of activated MEK1 in differentiating epidermal cells is sufficient to generate hyperproliferative and inflammatory skin lesions. *J. Invest. Dermatol.* **123**, 503–515 (2004).
- Murphy, J. E., Morales, R. E., Scott, J. & Kupper, T. S. IL-1 $\alpha$ , innate immunity, and skin carcinogenesis: the effect of constitutive expression of IL-1 $\alpha$  in epidermis on chemical carcinogenesis. *J. Immunol.* **170**, 5697–5703 (2003).
- Werner, S. & Smola, H. Paracrine regulation of keratinocyte proliferation and differentiation. *Trends Cell Biol.* **11**, 143–146 (2001).
- Hauser, C., Saurat, J. H., Schmitt, A., Jaunin, F. & Dayer, J. M. Interleukin 1 is present in normal human epidermis. *J. Immunol.* **136**, 3317–3323 (1986).
- Kaufman, C. K. & Fuchs, E. It's got you covered. NF- $\kappa$ B in the epidermis. *J. Cell Biol.* **149**, 999–1004 (2000).
- Murphy, J. E., Robert, C. & Kupper, T. S. Interleukin-1 and cutaneous inflammation: a crucial link between innate and acquired immunity. *J. Invest. Dermatol.* **114**, 602–608 (2000).
- Yamanaka, K. *et al.* Skin-specific caspase-1-transgenic mice show cutaneous apoptosis and pre-endotoxin shock condition with a high serum level of IL-18. *J. Immunol.* **165**, 997–1003 (2000).
- Groves, R. W., Mizutani, H., Kieffer, J. D. & Kupper, T. S. Inflammatory skin disease in transgenic mice that express high levels of interleukin 1 $\alpha$  in basal epidermis. *Proc. Natl Acad. Sci. USA* **92**, 11874–11878 (1995).
- Faustin, B. & Reed, J. C. Sunburned skin activates inflammasomes. *Trends Cell Biol.* **18**, 4–8 (2008).
- Feldmeyer, L. *et al.* The inflammasome mediates UVB-induced activation and secretion of interleukin-1 $\beta$  by keratinocytes. *Curr. Biol.* **17**, 1140–1145 (2007).
- Johansen, C., Moeller, K., Kragballe, K. & Iversen, L. The activity of caspase-1 is increased in lesional psoriatic epidermis. *J. Invest. Dermatol.* **127**, 2857–2864 (2007).
- Al-Mashat, H. A. *et al.* Diabetes enhances mRNA levels of proapoptotic genes and caspase activity, which contribute to impaired healing. *Diabetes* **55**, 487–495 (2006).
- Chun, H. J. *et al.* Pleiotropic defects in lymphocyte activation caused by caspase-8 mutations lead to human immunodeficiency. *Nature* **419**, 395–399 (2002).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank S. Hedrick, W. Havran, B. Yu, D. Witherden, A. Hoffmann, D. Stachura and members of the Jamora laboratory for providing reagents and helpful discussions. This work was supported by grants from the National Institutes of Health (NIAMS grant number 5R01AR053185-03) and the American Skin Association, and a Career Award from the Dermatology Foundation.

**Author Contributions** P.L., D.L., C.C., S.C. and C.J. performed the experiments; I.C. engineered the caspase 8 floxed mice; P.L. and C.J. designed the experiments; P.L., D.L. and C.J. wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to C.J. ([cjamora@ucsd.edu](mailto:cjamora@ucsd.edu)).

## METHODS

**Histology, *in situ* hybridization and immunohistochemistry.** Mouse skin isolated from P10 wild-type and knockout animals were either frozen in OCT (Tissue-Tek) or fixed overnight in Bouin's fixative and embedded in paraffin, depending on the application. Paraffin sections were prepared for *in situ* hybridization and histology, and counterstained with haematoxylin and eosin. *In situ* hybridization was performed using a digoxigenin-UTP (Roche) labelled probe against the 3'-untranslated region of mouse caspase 8 and the NALP3 probe was generated as described previously<sup>26</sup>. Staining for anti-cyclin D1 (Zymed), p-MAPK (Cell Signaling), p-NFκB (Ser276, Cell Signaling), (P)-p38 (Cell Signaling), caspase 1, p20 (M-19, Santa Cruz) and Ki67 (Vector Labs) was performed on 8-μm paraffin sections after antigen-retrieval. Epidermal differentiation markers K5, K1, loricrin, filaggrin and the wound healing marker K6 (Covance) were stained in 8-μm frozen sections after the tissues were fixed for 10 min in 4% paraformaldehyde. Immune cell infiltrates were stained using FITC-conjugated antibodies against MAC-1, GR-1, CD3, CD4, CD8, γδTCR and Vγ3 (BD Biosciences). For nuclear staining, Hoechst 33342 (Calbiochem) was added at a final concentration of 1 mg ml<sup>-1</sup> to the secondary antibody dilution. Immunofluorescence was detected using rhodamine-X or FITC-conjugated secondary antibodies (Jackson ImmunoResearch), or expression was developed using the Vectastain ABC kit (Vector Labs) according to the manufacturer's instructions.

**Cornified envelope assay.** Isolation of corneocytes in the squamous layer was performed as previously described<sup>27</sup>.

**Quantitative real-time PCR.** Total RNA was extracted from whole skin of wild-type ( $n = 3$ ) and knockout ( $n = 3$ ) mice at P10 or 7 days post-wounded skin ( $n = 3$ ) using Trizol reagent (Invitrogen) according to manufacturer's instructions. Similarly, epidermis from wild-type ( $n = 3$ ) and knockout ( $n = 3$ ) P3 mice was separated with dispase treatment for 1 h at 37 °C and total RNA was isolated using the Trizol protocol. cDNA was synthesized by reverse transcription using oligo-dT as primers (Superscript III kit, Invitrogen). Real-time PCR analysis was performed with previously described primers<sup>28,29</sup> using a Stratagene Mx-3000 system. Reactions were performed using the Brilliant SYBR Green QPCR reagent mix (Stratagene) and experiments were carried out in triplicate from cDNA isolated from six different animals. Data shown is a representative example. The specificity of reactions was determined by dissociation curve analysis and quantification analysis was performed using the Mx-3000 software.

**Proliferation assays.** Epidermis of newborn caspase 8 wild-type and knockout mice were separated in dispase overnight at 4 °C. Isolated epidermis or dermis was incubated in 3 ml of keratinocyte media<sup>29</sup> for 3 days. For Fig. 4, conditioned

media were diluted 1:5 with keratinocytes media and applied to a well containing 10,000 keratinocytes. Cells were collected by trypsinization every 24 h for 4 days and counted on a haemocytometer, and results are representative of three independent experiments performed in triplicate. To prime dermal fibroblasts, 400 μl of epidermal-conditioned media was diluted in 1.6 ml keratinocytes media and added to a confluent 6-cm dish of fibroblasts and incubated for 3 h at 37 °C. This epidermal-primed, dermal-fibroblast conditioned media was then diluted 1:2 in keratinocyte media and added to primary keratinocytes with 2 ng ml<sup>-1</sup> IL1α-neutralizing antibody or control antibody, or rhIL1α.

**NFκB translocation assay.** Diluted (1:5) wild-type and knockout epidermal and dermal conditioned media described earlier was applied to coverslips seeded with 30,000 primary keratinocytes 16 h before. Cells were incubated with this conditioned media and control goat antibody, 2 ng ml<sup>-1</sup> IL1α-neutralizing antibody, or 10 pg ml<sup>-1</sup> rhIL1α for 3 h at 37 °C. Cells were stained with an antibody recognizing NFκB (Santa Cruz)<sup>29</sup>.

**IL1α ELISA.** Isolated epidermis from P3 caspase 8 wild-type and knockout mice were incubated for 24 h in 2 ml of keratinocyte media containing DMSO or inhibitors of AKT (20 μM), p38 MAPK (SB 203580; 2.4 μM), caspase 1 (80 μM) or calpain (200 μM) (all from Calbiochem). Secreted mouse IL1α levels were detected using the Quantikine kit from Stratagene according to manufacturer's instructions.

**Caspase 1 activity assay.** Explants were treated as described earlier, and after 24 h of incubation with vehicle control (DMSO) or inhibitors they were trypsinized for 20 min at room temperature. Protein lysates were extracted from isolated epidermal cells and 100 μg of protein was analysed for caspase 1 activity as previously described<sup>30</sup>. Data shown is representative of three independent experiments performed in triplicate.

26. Anderson, J. P. *et al.* Structural, expression, and evolutionary analysis of mouse CIAS1. *Gene* **338**, 25–34 (2004).
27. Hohl, D. *et al.* Characterization of human loricrin. Structure and function of a new class of epidermal cell envelope proteins. *J. Biol. Chem.* **266**, 6626–6636 (1991).
28. Aravalli, R. N., Hu, S., Rowen, T. N., Palmquist, J. M. & Lokensgard, J. R. Cutting edge: TLR2-mediated proinflammatory cytokine and chemokine production by microglial cells in response to herpes simplex virus. *J. Immunol.* **175**, 4189–4193 (2005).
29. Kobiela, A. & Fuchs, E. Links between α-catenin, NF-κB, and squamous cell carcinoma in skin. *Proc. Natl Acad. Sci. USA* **103**, 2322–2327 (2006).
30. Li, J., Yin, H. L. & Yuan, J. Flightless-1 regulates proinflammatory caspases by selectively modulating intracellular localization and caspase activity. *J. Cell Biol.* **181**, 321–333 (2008).

## LETTERS

# Sphingosine-1-phosphate mobilizes osteoclast precursors and regulates bone homeostasis

Masaru Ishii<sup>1,4,5</sup>, Jackson G. Egen<sup>1</sup>, Frederick Klauschen<sup>2</sup>, Martin Meier-Schellersheim<sup>2</sup>, Yukihiro Saeki<sup>5</sup>, Jean Vacher<sup>6</sup>, Richard L. Proia<sup>3</sup> & Ronald N. Germain<sup>1,2</sup>

Osteoclasts are the only somatic cells with bone-resorbing capacity and, as such, they have a critical role not only in normal bone homeostasis (called 'bone remodelling') but also in the pathogenesis of bone destructive disorders such as rheumatoid arthritis and osteoporosis<sup>1</sup>. A major focus of research in the field has been on gene regulation by osteoclastogenic cytokines such as receptor activator of NF- $\kappa$ B-ligand (RANKL, also known as TNFSF11) and TNF- $\alpha$ , both of which have been well documented to contribute to osteoclast terminal differentiation<sup>2,3</sup>. A crucial process that has been less well studied is the trafficking of osteoclast precursors to and from the bone surface, where they undergo cell fusion to form the fully differentiated multinucleated cells that mediate bone resorption. Here we report that sphingosine-1-phosphate (S1P), a lipid mediator enriched in blood<sup>4,5</sup>, induces chemotaxis and regulates the migration of osteoclast precursors not only in culture but also *in vivo*, contributing to the dynamic control of bone mineral homeostasis. Cells with the properties of osteoclast precursors express functional S1P<sub>1</sub> receptors and exhibit positive chemotaxis along an S1P gradient *in vitro*. Intravital two-photon imaging of bone tissues showed that a potent S1P<sub>1</sub> agonist, SEW2871, stimulated motility of osteoclast precursor-containing monocytoid populations *in vivo*. Osteoclast/monocyte (CD11b, also known as ITGAM) lineage-specific conditional S1P<sub>1</sub> knockout mice showed osteoporotic changes due to increased osteoclast attachment to the bone surface. Furthermore, treatment with the S1P<sub>1</sub> agonist FTY720 relieved ovariectomy-induced osteoporosis in mice by reducing the number of mature osteoclasts attached to the bone surface. Together, these data provide evidence that S1P controls the migratory behaviour of osteoclast precursors, dynamically regulating bone mineral homeostasis, and identifies a critical control point in osteoclastogenesis that may have potential as a therapeutic target.

Because of its role in controlling cell migration in other tissues, we initially focused on S1P and the S1P family of receptors (previously known as EDG receptors) responsive to this lipid mediator. The murine monocyte cell line RAW264.7, and mouse bone-marrow-derived M-CSF (colony stimulating factor-1, also known as CSF1)-dependent monocytes (termed BM-MDM), which can both differentiate into osteoclast-like cells on exposure to RANKL and represent widely used *in vitro* models of osteoclast precursors, each express messenger RNAs encoding two receptors for S1P (S1P<sub>1</sub> and S1P<sub>2</sub>, encoded by *S1PR1* and *S1PR2*, respectively). The expression of *S1PR1* mRNA decreased after RANKL stimulation (Fig. 1a, b), and this transcriptional repression was dependent on NF- $\kappa$ B (blocked by BAY11-7085) but not on NF-AT (also known as NFATC1; blocked by cyclosporin A) (Fig. 1b), both of which have been recognized as

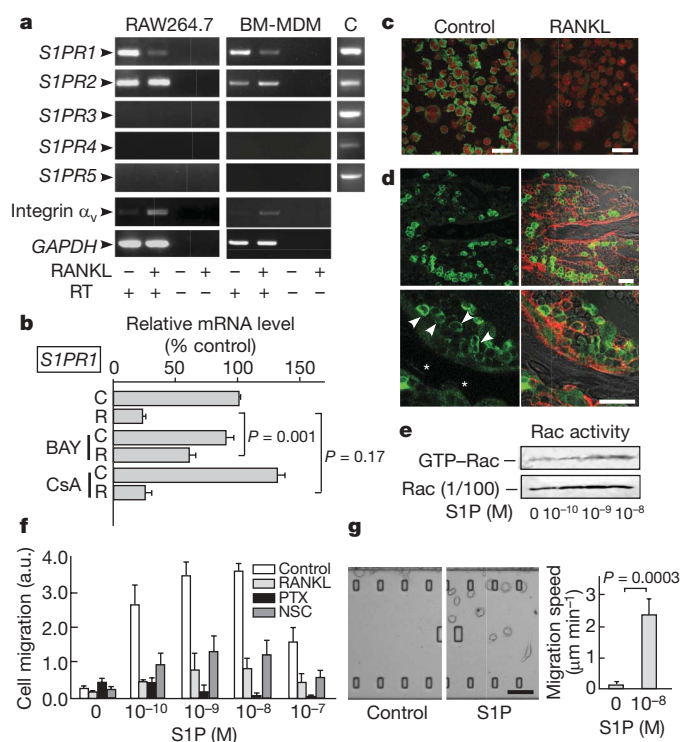
crucial transcriptional factors for osteoclastogenesis<sup>1-3</sup>. S1P<sub>1</sub> protein expression, as detected by immunocytochemistry, was also reduced by RANKL stimulation (Fig. 1c). In accord with these *in vitro* findings, strong immunoreactivity for S1P<sub>1</sub> could be detected in mononuclear cells in native bone tissues (Fig. 1d, arrowheads), whereas weaker staining was seen in large cells lining bone trabeculae (Fig. 1d, asterisks) that are probably mature osteoclasts (mean fluorescence intensities:  $49.1 \pm 5.8$  and  $19.4 \pm 2.0$  in mononuclear cells and attached cells, respectively;  $P = 0.0003$ ). Both RAW264.7 (Fig. 1f) and BM-MDM (data not shown) cells showed positive chemotactic responses to an S1P gradient *in vitro*, and these responses were greatly attenuated by pretreatment with RANKL (Fig. 1f). This latter finding is concordant with the RANKL-dependent reduction in *S1PR1* gene expression (Fig. 1a). S1P-directed chemotaxis was blocked by pertussis toxin (PTX)<sup>6</sup> and NSC23766 (ref. 7) (Fig. 1f), and S1P induced an increase in the GTP-bound, active form of Rac (GTP-Rac) (Fig. 1e), suggesting that Rac and the adenylyl-cyclase-inhibiting G protein G $\alpha_i$  are involved in the S1P<sub>1</sub> chemotactic signalling pathway in these cells, consistent with previous reports<sup>8,9</sup>. *In vitro* chemotaxis of RAW264.7 was also examined using a dynamic imaging approach, in which cells could be observed vigorously migrating towards S1P-containing medium (Fig. 1g and Supplementary Videos 1 and 2). We also detected S1P-induced upregulation of integrins expressed in monocytoid cells, suggesting that S1P-mediated alterations in cell adhesion could contribute to chemotaxis *in vivo* (Supplementary Fig. 1).

To investigate whether osteoclast precursor migration is affected by S1P gradients *in vivo*, we performed intravital two-photon imaging of calvaria bone tissues<sup>10,11</sup> and examined the migratory behaviour of monocytoid cells containing osteoclast precursors resident in the marrow spaces. We used two strains of mice expressing enhanced green fluorescent protein (EGFP) in subsets of myeloid cells: CX<sub>3</sub>CR1-EGFP knock-in (heterozygous) mice<sup>12,13</sup> and CSF1R-EGFP transgenic mice<sup>14</sup>. In both strains, haematopoietic monocytoid lineage cell types predominantly expressed EGFP<sup>12,13</sup> and we confirmed that tartrate-resistant acid phosphatase (TRAP)-positive mature osteoclasts expressed EGFP in these animals (Supplementary Fig. 2), suggesting that EGFP<sup>+</sup> cells contain osteoclast precursor monocytes. Nearly two-thirds of CX<sub>3</sub>CR1-EGFP cells and half of CSF1R-EGFP cells express RANK (also known as TNFRSF11A), a receptor for RANKL (Supplementary Table 1), suggesting that there is a capacity for many of the fluorescent cells to differentiate into mature osteoclasts.

CX<sub>3</sub>CR1-EGFP-positive cells present in bone marrow stromal locations or at the bone surface and not in the blood sinusoidal spaces were generally stationary under control conditions (Fig. 2a, upper panels

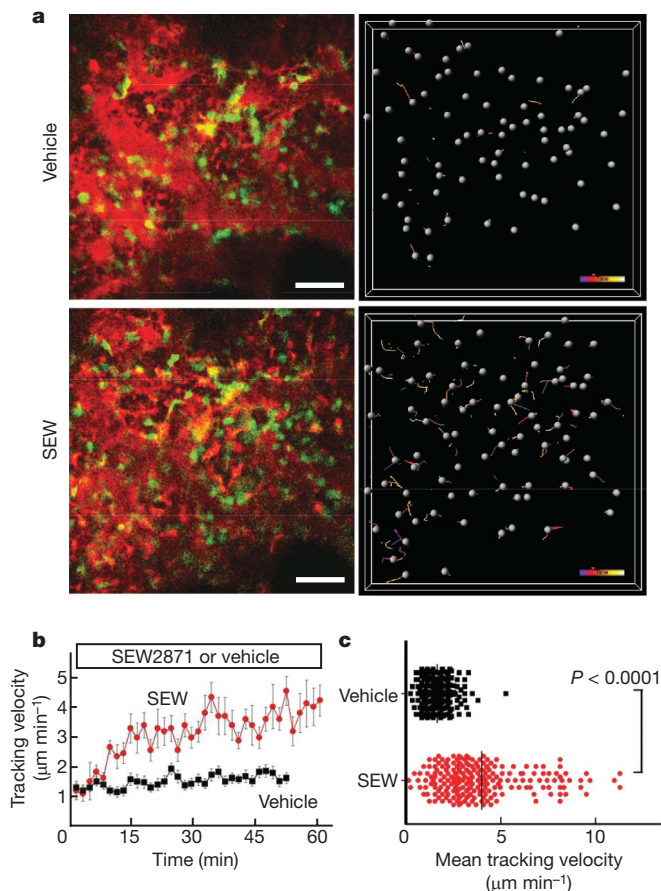
<sup>1</sup>Lymphocyte Biology Section, Laboratory of Immunology, <sup>2</sup>Program in Systems Immunology and Infectious Disease Modeling, National Institute of Allergy and Infectious Diseases, and <sup>3</sup>Genetics of Development and Disease Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-1892, USA. <sup>4</sup>Laboratory of Biological Imaging, WPI-Immunology Frontier Research Center, Osaka University, Osaka 565-0871, Japan. <sup>5</sup>Department of Clinical Research, National Osaka Minami Medical Center, Osaka 586-8521, Japan. <sup>6</sup>Institut de Recherches Cliniques de Montréal, Québec H2W 1R7, Canada.





**Figure 1 | Expression and function of S1P receptors in osteoclast precursor monocytes.** **a**, Expression of mRNAs encoding 5 mammalian S1P receptors (S1P<sub>1</sub> to S1P<sub>5</sub>, encoded by *S1PR1* to *S1PR5*) in RAW264.7 monocytes (left panels) and in mouse BM-MDM (middle panels), detected by RT-PCR. mRNAs for integrin  $\alpha_v$  and *GAPDH* were also analysed as controls. Total complementary DNA isolated from mouse thymus was used for positive controls (C, right panels). RT, reverse transcription. **b**, qRT-PCR analysis of *S1PR1* mRNA expressed in RAW264.7 cells cultured in the absence (C) or presence (R) of RANKL. Cells were treated with BAY11-7085 (BAY; 10  $\mu$ M) or cyclosporine A (CsA; 1  $\mu$ M). Error bars represent s.e.m. **c**, Immunofluorescent detection of S1P<sub>1</sub> protein (green) in RAW264.7 cells cultured in the absence (left panel) or presence (right panel) of 50 ng ml<sup>-1</sup> RANKL. Nuclei were visualized with propidium iodide (red). **d**, Immunohistochemical analysis of S1P<sub>1</sub> in mouse femoral bone tissues at low (top) and high (bottom) magnification. Staining for S1P<sub>1</sub> (green; left two panels) and a merged image with staining for CD9 (red) and transmission (Nomarski image) (right two panels) are shown. Arrowheads represent S1P<sub>1</sub><sup>high</sup> mononuclear cells adjacent to bone trabeculae (asterisk). Scale bars in **c** and **d**, 20  $\mu$ m. **e**, S1P-induced Rac stimulation. RAW264.7 cells were treated with various concentrations of S1P for 15 min and then analysed for GTP-Rac. **f**, *In vitro* chemotactic response of RAW264.7 to S1P gradient. Error bars represent s.e.m. ( $n = 6$ ). **g**, *In vitro* S1P-directed chemotaxis of RAW264.7 dynamically visualized using EZ-Taxiscan. Cells were loaded onto the upper chamber and the lower chamber was filled either with normal medium (right panel; Supplementary Video 1) or with medium containing 10<sup>-8</sup> M S1P (middle panel; Supplementary Video 2). Mean migration speed is shown in right panel. Error bars represent s.e.m. ( $n = 8$ ). Scale bar represents 100  $\mu$ m.

and Supplementary Video 3). A subset of the labelled cells became motile shortly after the intravenous application of SEW2871 (refs 15 and 16; 5 mg kg<sup>-1</sup>), a selective agonist for the S1P<sub>1</sub> receptor (Fig. 2a, lower panels, and Supplementary Video 4), with some of the mobilized cells entering the blood circulation (data not shown). A similar but less pronounced effect of SEW2871 on myeloid cell motility was observed in the bone marrow of CSF1R-EGFP transgenic mice (Supplementary Fig. 3 and Supplementary Videos 5 and 6). The difference in the fraction of cells mobilized by SEW2871 treatment in the two mouse strains is probably due to the fact that the EGFP-expressing subsets in these animals are not completely overlapping (Supplementary Table 1). Consistent with these findings, we also observed an increase in the percentage and absolute number of



**Figure 2 | *In vivo* S1P-mediated increase in motility of osteoclast precursor monocytes visualized using intravital two-photon imaging.** **a**, Intravital two-photon imaging of mouse skull bone tissues of heterozygous CX<sub>3</sub>CR1-EGFP knock-in mice, in the absence (top panels; Supplementary Video 3) or presence (bottom panels; Supplementary Video 4) of the S1P<sub>1</sub> agonist SEW2871 (SEW; 5 mg kg<sup>-1</sup>). CX<sub>3</sub>CR1-EGFP-positive cells appear green. The microvasculature was visualized by intravenous injection of 70 kDa dextran-conjugated Texas Red (red; left panels). The movements of CX<sub>3</sub>CR1-EGFP positive cells were tracked for 10 min (right panels). Grey spheres represent cells and coloured lines show the associated trajectories. Scale bars represent 50  $\mu$ m. **b**, Quantification of CX<sub>3</sub>CR1-EGFP-positive cell velocity. Tracking velocity over time after application of SEW2871 (red circle) or vehicle only (black square) are shown. Data points represent mean values  $\pm$  s.e.m. of cell velocities in the field at certain time points ( $n = 15$  for vehicle and  $n = 14$  for SEW2871). **c**, Summary of mean velocity of CX<sub>3</sub>CR1-EGFP positive cells treated with SEW2871 (red circle) or vehicle (black circle). Data points ( $n = 231$  for vehicle and  $n = 210$  for SEW2871) represent individual cells compiled from five independent experiments.

monocytoid cells in peripheral blood, and their reciprocal decrease in bone marrow but not spleen or liver, after SEW2871 treatment as compared to vehicle controls (Supplementary Fig. 4 and Supplementary Table 2). These results suggest that bone-marrow-resident monocytes, including many osteoclast precursors, become motile and exit marrow spaces in response to S1P stimulation.

To evaluate the *in vivo* effect of S1P-directed chemotaxis of osteoclast precursors on bone homeostasis, we examined mice deficient in S1P<sub>1</sub>. Because S1P<sub>1</sub> deficiency causes embryonic lethality<sup>17</sup>, we generated osteoclast/monocyte-specific S1P<sub>1</sub>-deficient (*S1PR1*<sup>-/-</sup>) mice by crossing animals bearing conditional S1P<sub>1</sub> knockout alleles (*S1PR1*<sup>loxP</sup>)<sup>18</sup> to transgenic mice expressing Cre under the *CD11b* promoter<sup>19</sup>, which facilitates deletion of *loxP*-flanked sequences in monocyte-lineage myeloid cells including osteoclasts. *S1PR1* mRNA was essentially undetectable by conventional and real-time quantitative PCR with reverse transcription (qRT-PCR) analyses in *CD11b*<sup>+</sup> myeloid cells purified from the *S1PR1*<sup>-/-</sup> mice

(Supplementary Fig. 5a, b) and immunoreactivity for S1P<sub>1</sub> was greatly diminished in *S1PR1*<sup>-/-</sup> bone tissues (Supplementary Fig. 5c), indicating that there was almost complete ablation of S1P<sub>1</sub> expression in this lineage.

Morphohistometric analyses using micro-CT ( $\mu$ CT) showed that femora of mice genotyped as *S1PR1*<sup>loxP/loxP</sup> *CD11b-Cre* (*S1PR1*<sup>-/-</sup>) were osteoporotic, compared with those of control (*S1PR1*<sup>+/+</sup> *CD11b-Cre*) littermates (Fig. 3a, upper panels). Bone tissue density (bone volume/total volume) of *S1PR1*<sup>-/-</sup> mice was significantly lower than that of controls (Fig. 3b), and both trabecular thickness and trabecular density were decreased in *S1PR1*<sup>-/-</sup> bones (Fig. 3c). Mice genotyped as *S1PR1*<sup>loxP/+</sup> *CD11b-Cre* (heterozygous) were indistinguishable from controls, indicating that a single *S1PR1* allele is sufficient to sustain normal S1P<sub>1</sub>-mediated functions.

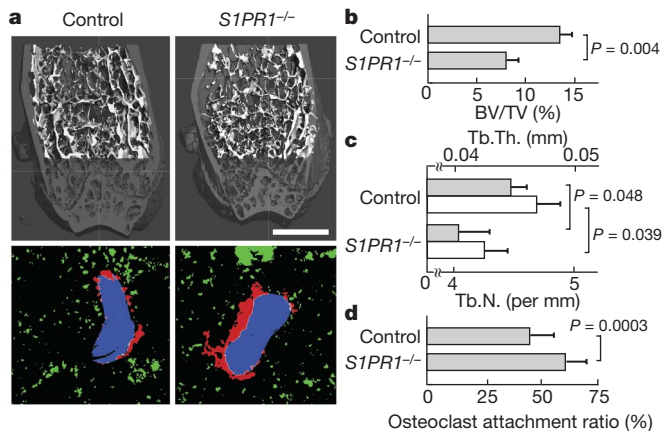
Bone surface attachment of mature osteoclasts was evaluated by immunohistological examination. Both osteoclasts and bone matrices were visualized simultaneously by fluorescence-based staining for TRAP<sup>20</sup> and by collagen-derived second harmonic signals elicited by two-photon excitation<sup>21</sup>, respectively (Supplementary Fig. 6). Images were automatically segmented off-line using a custom software package (F.K., submitted manuscript). An 'osteoclast attachment ratio' (bone surface occupied by osteoclasts/total bone surface) was calculated automatically without manual intervention (Fig. 3a, lower panels and Supplementary Fig. 6). The computational analyses showed that osteoclast attachment to the bone surface was significantly enhanced in *S1PR1*<sup>-/-</sup> animals, as compared to controls (Fig. 3d), consistent with the decrease in bone density in *S1PR1*<sup>-/-</sup> mice (Fig. 3b, c). Because S1P has no measurable effect on *in vitro* osteoclast differentiation stimulated by RANKL (Supplementary Fig. 7), a reasonable interpretation of these data are that the loss of S1P<sub>1</sub> function alters the chemotactic behaviour of osteoclast precursors rather than their terminal maturation to osteoclasts or the bone-resorbing action of osteoclasts.

Because the concentration of S1P in blood is higher than in tissues, S1P-mediated chemotaxis of osteoclast precursors would thus be

expected to contribute to their recirculation from bone tissues to systemic blood flow, limiting formation of mature osteoclasts and reducing bone destruction. The present result showing the increased deposition of *S1PR1*<sup>-/-</sup> osteoclasts on the bone surface strongly supports the concept of S1P-driven precursor recirculation. It has already been reported that osteoclast precursor monocytes (RAW264.7) show positive chemotaxis *in vitro* towards CXCL12 (also known as SDF-1), a chemokine secreted from bone marrow stromal cells<sup>22</sup>, leading to a straightforward competitive model for the migration and positioning of osteoclasts (Supplementary Fig. 8). Monocyte-lineage osteoclast precursors come close to the bone surface attracted by CXCL12, however not all of the recruited cells undergo differentiation, and some of the precursors recirculate into the blood flow owing to attraction by S1P. Some osteoclast precursors may be directly derived within bone marrow spaces (Supplementary Fig. 8, dashed arrow), and they are also susceptible to S1P-driven 'circulation' into the systemic flow. During osteoclast differentiation induced by RANKL, the expression of S1P<sub>1</sub> is repressed (Fig. 1a) which ensures stable localization of the maturing cells on the bone surface. Regulation of osteoclast generation by competing chemokine pathways that attract precursors to and from the bone surface allows finely-tuned dynamic control of osteoclastogenesis during health and in various conditions such as inflammation and metabolic disorders. Bone surface CXCL12 expression has recently been shown to be regulated by sympathetic neuron stimulation<sup>23</sup>, suggesting that there is an interesting interaction between circulating S1P and the autonomic nervous system for the regulation of osteoclast precursor mobilization.

This newly revealed role of S1P-directed recirculation of osteoclast precursor in bone homeostasis prompted us to examine possible therapeutic implications of this regulatory pathway. FTY720 (Fingolimod) is metabolized after administration to an agonist for four of the five S1P receptors, including S1P<sub>1</sub> (but not S1P<sub>2</sub>), and has emerged as a clinically promising immunosuppressive drug<sup>24,25</sup> that presumably acts by limiting effector lymphocyte egress from lymph nodes. FTY720 is at present being tested clinically in multiple sclerosis and allogeneic transplantation<sup>26,27</sup>, but has not been examined for its effect on bone homeostasis. Here we show that FTY720 has a capacity to ameliorate bone loss in a model of postmenopausal osteoporosis.

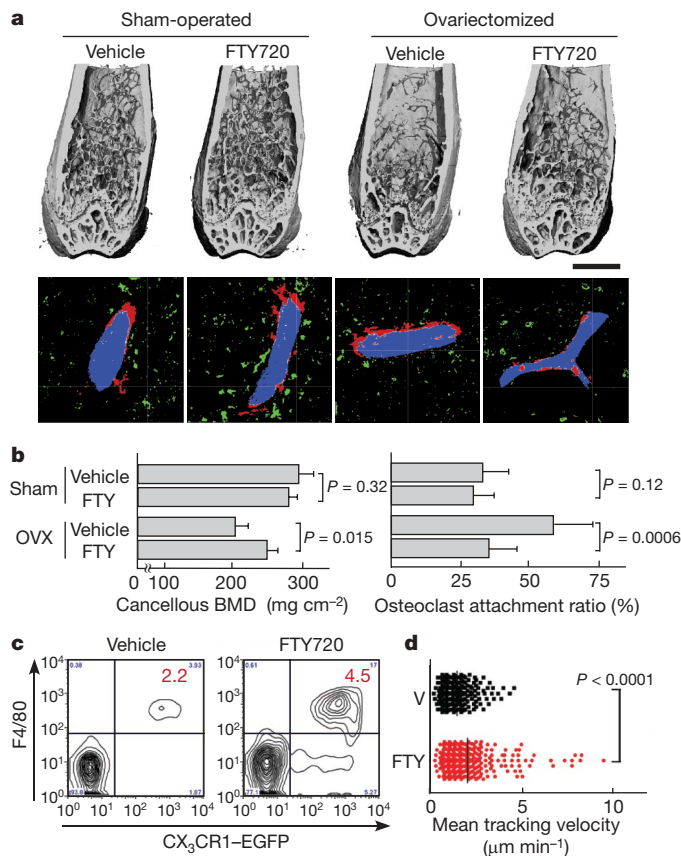
Either ovariectomized or sham-operated mice were administered FTY720 (3 mg kg<sup>-1</sup> day<sup>-1</sup>) for 4 weeks after which excised femora were analysed by  $\mu$ CT (Fig. 4a, b). FTY720 treatment significantly prevented bone density loss after ovariectomy, but had no effects under control (sham-operated) conditions. Computational morphometrical analyses showed that osteoclast attachment was increased after ovariectomy and partially restored to normal levels by FTY720 administration. We confirmed that FTY720 had no effect on oestrogen levels after ovariectomy (Supplementary Table 3). These results indicate that FTY720 relieves ovariectomy-induced bone loss by reducing osteoclast deposition onto bone surfaces. This effect correlated with an increase in circulating CX<sub>3</sub>CR1-EGFP<sup>+</sup> F4/80<sup>+</sup> (Fig. 4c) and CD11b<sup>+</sup> monocytes (Supplementary Fig. 9), both of which populations include osteoclast precursors, and a concomitant decrease in monocytoïd cells including osteoclast precursors in the bone marrow (Supplementary Fig. 9 and Supplementary Table 2). In addition, intravital two-photon imaging showed that FTY720 significantly stimulated motility of CX<sub>3</sub>CR1-EGFP<sup>+</sup> monocyte-lineage cells (Fig. 4d and Supplementary Video 7). The short-term mobility change seen using FTY720 (as observed by intravital microscopy) was less prominent than that obtained with SEW2871 (Fig. 2). This difference may arise because SEW2871 is active after intravenous injection, whereas FTY720 needs to be metabolized (phosphorylated) before becoming functional (Supplementary Fig. 10). We cannot rule out, however, the possibility that phospho-FTY720 can activate S1P receptors other than S1P<sub>1</sub> that are expressed on other cell types, such as endothelial cells. FTY720 was shown to have no effects on the expression of RANKL or osteoprotegerin (data not shown). Taken together, these results



**Figure 3 | In vivo effect of S1P<sub>1</sub> on bone mineral metabolism.**

**a**, Morphohistometric analyses of control and *S1PR1*<sup>-/-</sup> (*S1PR1*<sup>loxP/loxP</sup> *CD11b-Cre*) littermates. Femoral bone samples were analysed by cone-beam  $\mu$ CT (top panels) and by histological examination combined with computational quantification for measuring the osteoclast attachment ratio to the bone surface (bottom panels; see also Supplementary Fig. 6 and Supplementary Information). Blue areas represent bone trabeculae (second harmonic fluorescence signal), red and green areas show TRAP-positive osteoclasts that are attached to or detached from bone trabeculae, respectively, and white areas show the area of osteoclast/bone attachment. Scale bar represents 1 mm. **b–d**, Summary of the data of (b) bone matrix density (bone volume/total volume, BV/TV), (c) trabecular thickness (Tb.Th; filled bars) and trabecular density (Tb.N; open bars) calculated from  $\mu$ CT images, and (d) osteoclast attachment ratio calculated by computational segmentation analyses. Error bars represent s.e.m.,  $n = 3$  (b, c) and  $n = 40$  (from 3 mice) (d) for each.





**Figure 4 | Preventive effect of FTY720 on ovariectomy-induced osteoporosis.** **a**, Effect of FTY720 on bone mineral metabolism. Femurs were collected from mice after four different treatments: sham-operated and vehicle treated, sham-operated and FTY720 treated, ovariectomized and vehicle treated, and ovariectomized and FTY720 treated. Bone samples were analysed by cone-beam  $\mu$ CT (top panels) and by histological examination combined with computational quantification for measuring the osteoclast attachment ratio to the bone surface (bottom panels; see Supplementary Fig. 5). Scale bar represents 1 mm. **b**, Summary of the data of cancellous bone mineral density calculated from  $\mu$ CT images (left panel) and of osteoclast attachment ratio (right panel). Error bars represent s.e.m.,  $n = 3$  (left panel) and  $n = 20$  (from 3 mice; right panel) for each. **c**, Effect of FTY720 on the composition of PMCs. PMCs collected from CX<sub>3</sub>CR1-EGFP knock-in mice administered vehicle or FTY720 were stained with anti-F4/80 (Alexa647). Absolute numbers of CX<sub>3</sub>CR1-EGFP<sup>+</sup> F4/80<sup>+</sup> cells (per 10<sup>5</sup> ml of peripheral blood) are described in the panels. **d**, Effect of FTY720 on the mobility of CX<sub>3</sub>CR1-EGFP marked cells. Summary of mean tracking velocity of CX<sub>3</sub>CR1-EGFP-positive cells treated with FTY720 (FTY, red circles) or vehicle (V, black circles). Data points ( $n = 246$  for vehicle and  $n = 339$  for FTY720) represent individual cells compiled from four independent experiments. Intravital two-photon images of mouse skull bone tissues of heterozygous CX<sub>3</sub>CR1-EGFP knock-in mice treated with FTY720 are shown in Supplementary Video 7.

indicate that FTY720, a potent S1P<sub>1</sub> agonist, promotes the recirculation of osteoclast precursor monocytes from bone surface, an effect that counteracts ovariectomy-induced augmentation of osteoclastogenesis. This agonist effect of FTY720 on myeloid cells directly accessible to the blood circulation differs from the consequence of lymphocyte S1P<sub>1</sub> engagement by phosphorylated FTY720 within secondary lymphoid tissues, in which case downmodulation of the receptor results in functional antagonism of this pathway.

The present study provides evidence that S1P, as a circulation-attraction molecule, controls the migratory behaviour of osteoclast precursors, dynamically regulating bone mineral homeostasis in collaboration with bone-attraction molecules, such as CXCL12, and identifies a critical control point in osteoclastogenesis that has potential as a therapeutic target.

## METHODS SUMMARY

The RAW264.7 cell line and mouse BM-MDM cells, which contains osteoclast precursor cells, were cultured and stimulated to differentiate using RANKL as previously described<sup>28,29</sup>. *In vitro* chemotactic activity was evaluated using a modified Boyden's chamber<sup>30</sup> or with a visually-accessible chemotactic chamber, EZ-Taxiscan. *In vivo* S1P-directed chemotaxis of osteoclast precursor monocytes was visualized by intravital two-photon microscopy of mouse calvaria bone tissues of heterozygous CX<sub>3</sub>CR1-EGFP knock-in mice<sup>12</sup> and of CSF1R-EGFP transgenic<sup>14</sup> mice, according to a protocol modified from a previous report<sup>10,11</sup>. The generation of the loxP-flanked *S1PR1* allele (*S1PR1*<sup>loxP</sup>)<sup>18</sup> and osteoclast/monocyte lineage-specific *CD11b-Cre* transgenic mice<sup>19</sup> have been described previously. Histomorphometry of femurs from osteoclast/monocyte-lineage specific S1P<sub>1</sub> conditional knockout (*S1PR1*<sup>-/-</sup>) mice, as well as mice ovariectomized/sham-operated and treated with FTY720 or vehicle, were performed using  $\mu$ CT. Assessment of osteoclast attachment to the bone surface was performed using a newly developed fully automated segmentation approach to analyse two-photon images of bone tissue sections in which osteoclasts and bone trabeculae were visualized with fluorescent TRAP staining<sup>20</sup> and second harmonic generation from collagen fibres<sup>21</sup>, respectively. To examine the composition of peripheral blood mononuclear cells (PMCs), the cells were stained with fluorophore-conjugated anti-F4/80, anti-CD11b and anti-CD3 using conventional methods. Flow cytometric data were collected on a FACSCalibur and analysed with FlowJo software. All mice were bred and maintained under specific pathogen-free conditions at the animal facility of the National Institutes of Health (NIH), according to NIH institutional guidelines under an approved protocol. For statistical analyses, the Mann-Whitney rank sum test was used to calculate *P* values for highly skewed distributions. For Gaussian-like distributions, two-tailed *t*-tests were used.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 3 September; accepted 9 December 2008.

Published online 8 February 2009.

- Teitelbaum, S. L. Bone resorption by osteoclasts. *Science* **289**, 1504–1508 (2000).
- Teitelbaum, S. L. & Ross, F. P. Genetic regulation of osteoclast development and function. *Nature Rev. Genet.* **4**, 638–649 (2003).
- Karsenty, G. & Wagner, E. F. Reaching a genetic and molecular understanding of skeletal development. *Dev. Cell* **2**, 389–406 (2002).
- Rosen, H. & Goetzl, E. J. Sphingosine 1-phosphate and its receptors: an autocrine and paracrine network. *Nature Rev. Immunol.* **5**, 560–570 (2005).
- Cyster, J. G. Chemokines, sphingosine-1-phosphate, and cell migration in secondary lymphoid organs. *Annu. Rev. Immunol.* **23**, 127–159 (2005).
- Bokoch, G. M., Katada, T., Northup, J. K., Ui, M. & Gilman, A. G. Purification and properties of the inhibitory guanine nucleotide-binding regulatory component of adenylate cyclase. *J. Biol. Chem.* **259**, 3560–3567 (1984).
- Akbar, H., Cancelas, J., Williams, D. A., Zheng, J. & Zheng, Y. Rational design and applications of a Rac GTPase-specific small molecule inhibitor. *Methods Enzymol.* **406**, 554–565 (2006).
- Spiegel, S., English, D. & Milstien, S. Sphingosine 1-phosphate signaling: providing cells with a sense of direction. *Trends Cell Biol.* **12**, 236–242 (2002).
- Takuwa, Y. Subtype-specific differential regulation of Rho family G proteins and cell migration by the Edg family sphingosine-1-phosphate receptors. *Biochim. Biophys. Acta* **1582**, 112–120 (2002).
- Mazo, I. B. *et al.* Hematopoietic progenitor cell rolling in bone marrow microvessels: parallel contributions by endothelial selectins and vascular cell adhesion molecule 1. *J. Exp. Med.* **188**, 465–474 (1998).
- Mazo, I. B. *et al.* Bone marrow is a major reservoir and site of recruitment for central memory CD8<sup>+</sup> T cells. *Immunity* **22**, 259–270 (2005).
- Jung, S. *et al.* Analysis of fractalkine receptor CX<sub>3</sub>CR1 function by targeted deletion and green fluorescent protein reporter gene insertion. *Mol. Cell. Biol.* **20**, 4106–4114 (2000).
- Niess, J. H. *et al.* CX<sub>3</sub>CR1-mediated dendritic cell access to the intestinal lumen and bacterial clearance. *Science* **307**, 254–258 (2005).
- Burnett, S. H. *et al.* Conditional macrophage ablation in transgenic mice expressing a Fas-based suicide gene. *J. Leukoc. Biol.* **75**, 612–623 (2004).
- Sanna, M. G. *et al.* Sphingosine 1-phosphate (S1P) receptor subtypes S1P<sub>1</sub> and S1P<sub>2</sub>, respectively, regulate lymphocyte recirculation and heart rate. *J. Biol. Chem.* **279**, 13839–13848 (2004).
- Wei, S. H. *et al.* Sphingosine 1-phosphate type 1 receptor agonism inhibits transendothelial migration of medullary T cells to lymphatic sinuses. *Nature Immunol.* **6**, 1228–1235 (2005).
- Liu, Y. *et al.* Edg-1, the G protein-coupled receptor for sphingosine-1-phosphate, is essential for vascular maturation. *J. Clin. Invest.* **106**, 951–961 (2000).
- Allende, M. L. *et al.* G-protein-coupled receptor S1P<sub>1</sub> acts within endothelial cells to regulate vascular maturation. *Blood* **102**, 3665–3667 (2003).
- Ferron, M. & Vacher, J. Targeted expression of Cre recombinase in macrophages and osteoclasts in transgenic mice. *Genesis* **41**, 138–145 (2005).



20. Filgueira, L. Fluorescence-based staining for tartrate-resistant acidic phosphatase (TRAP) in osteoclasts combined with other fluorescent dyes and protocols. *J. Histochem. Cytochem.* **52**, 411–414 (2004).
21. Stoller, P., Reiser, K. M., Celliers, P. M. & Rubenchik, A. M. Polarization-modulated second harmonic generation in collagen. *Biophys. J.* **82**, 3330–3342 (2002).
22. Yu, X., Huang, Y., Collin-Osdoby, P. & Osdoby, P. Stromal cell-derived factor-1 (SDF-1) recruits osteoclast precursors by inducing chemotaxis, matrix metalloproteinase-9 (MMP-9) activity, and collagen transmigration. *J. Bone Miner. Res.* **18**, 1404–1418 (2003).
23. Katayama, Y. *et al.* Signals from the sympathetic nervous system regulate hematopoietic stem cell egress from bone marrow. *Cell* **124**, 407–421 (2006).
24. Matloubian, M. *et al.* Lymphocyte egress from thymus and peripheral lymphoid organs is dependent on S1P receptor 1. *Nature* **427**, 355–360 (2004).
25. Cyster, J. G. Chemokines, sphingosine-1-phosphate, and cell migration in secondary lymphoid organs. *Annu. Rev. Immunol.* **23**, 127–159 (2005).
26. Kappos, L. *et al.* Oral fingolimod (FTY720) for relapsing multiple sclerosis. *N. Engl. J. Med.* **355**, 1124–1140 (2006).
27. Tedesco-Silva, H. *et al.* FTY720 versus mycophenolate mofetil in de novo renal transplantation: six-month results of a double-blind study. *Transplantation* **84**, 885–892 (2007).
28. Kobayashi, K. *et al.* Tumor necrosis factor  $\alpha$  stimulates osteoclast differentiation by a mechanism independent of the ODF/RANKL-RANK interaction. *J. Exp. Med.* **191**, 275–286 (2000).
29. Ishii, M. *et al.* RANKL-induced expression of tetraspanin CD9 in lipid raft membrane microdomain is essential for cell fusion during osteoclastogenesis. *J. Bone Miner. Res.* **21**, 965–976 (2006).
30. Okamoto, H. *et al.* Inhibitory regulation of Rac activation, membrane ruffling, and cell migration by the G protein-coupled sphingosine-1-phosphate receptor EDG5 but not EDG1 or EDG3. *Mol. Cell. Biol.* **20**, 9247–9261 (2000).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank U. H. von Andrian and I. B. Mazo for their help with the technique of intravital skull bone imaging. We also thank Y. Takuwa and N. Sugimoto for discussions, and P. M. Murphy and S. Venkatesan for their help in imaging *in vitro* chemotaxis using the EZ-Taxiscan. This work was supported in part by the Intramural Research Program of the National Institute of Allergy and Infectious Diseases, NIH, US Department of Health and Human Services, and by a fellowship grant to M.I. from the International Human Frontier Science Program.

**Author Contributions** M.I. performed most of experiments, with the assistance of J.G.E. and Y.S. for two-photon microscopy and for the *in vitro* osteoclast culture system, respectively. F.K. developed the unsupervised segmentation software and performed the computational analyses used to quantify the osteoclast-bone surface interface, with the assistance of M.M.-S. J.V. and R.L.P. generated *CD11b-Cre* transgenic and *S1PR1<sup>loxP</sup>* knock-in mice, respectively. R.N.G. helped M.I. in designing and interpreting experiments, as well as in writing the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to R.N.G. ([rgermain@niaid.nih.gov](mailto:rgermain@niaid.nih.gov)).

## METHODS

**Cell culture.** The RAW264.7 cell line and mouse BM-MDM cells, which contain osteoclast precursor cells, were cultured as previously described<sup>28,29</sup>. To stimulate osteoclastogenesis, 50 ng ml<sup>-1</sup> RANKL (PeproTech) was added to the medium and the cells were incubated for 3 to 4 days.

**Conventional and quantitative RT-PCR amplification.** Conventional RT-PCR was performed as described previously<sup>29</sup>. Primers used are listed in Supplementary Table 4. Relative quantification with real-time RT-PCR was performed using an ABI PRISM 7900 (Applied Biosystems Inc.) with an Assay-on-Demand TaqMan probe and relevant primers, according to the manufacturer's instructions.

**Determination of Rac activity.** Pulldown assay methods to determine the GTP-bound active form of Rac have been described in detail previously<sup>30</sup>.

**In vitro chemotaxis chamber assay.** Chemotactic migration of cells was measured in a modified Boyden chamber as described previously<sup>27</sup>.

**EZ-Taxiscan chemotaxis assay.** Chemotaxis experiments were also conducted in an EZ-Taxiscan chamber according to the manufacturer's protocol (Effector Cell Institute). The EZ-Taxiscan is a visually-accessible chemotactic chamber, in which one compartment containing ligand (S1P) and another compartment containing cells are connected by a microchannel. Phase contrast images of migrating cells were acquired at 1 min intervals.

**Mice.** C57BL/6 mice, CX<sub>3</sub>CR1-EGFP knock-in mice<sup>12</sup>, CSF1R (M-CSF receptor)-EGFP transgenic mice<sup>14</sup>, were obtained from The Jackson Laboratory. The generation of the loxP-flanked *S1PR1* allele (*S1PR1<sup>loxP</sup>*)<sup>18</sup> and osteoclast/monocyte lineage-specific *CD11b-Cre* transgenic mice<sup>19</sup> were described previously. *S1PR1<sup>+/+</sup> × CD11b-Cre* and *S1PR1<sup>loxP/+</sup> × CD11b-Cre* littermates were born at expected frequencies and were viable and fertile, but live *S1PR1<sup>loxP/loxP</sup> × CD11b-Cre* littermates were obtained rarely (three viable *S1PR1<sup>loxP/loxP</sup> × CD11b-Cre* out of 59 total mice), suggesting possible embryonic lethality as seen with global S1P<sub>1</sub> deficiency. All mice were bred and maintained under specific pathogen-free conditions at animal facilities of NIH. Mutant mice were genotyped by PCR<sup>18,19</sup>. All mice were housed and handled according to NIH institutional guidelines under an approved protocol.

**Immunohistochemistry.** Immunohistological analyses were performed as described previously<sup>2</sup>. Anti-mouse S1P<sub>1</sub> rabbit polyclonal antibody was purchased from Affinity Bioreagents. In some experiments, fluorescence-based staining for TRAP<sup>20</sup> with ELF 97 substrate (Molecular Probes) was used with some modifications.

**Two-photon intravital bone tissue imaging.** Intravital microscopy of mouse calvaria bone tissues was performed using a protocol modified from a previous report<sup>10,11</sup>. Mice were anaesthetized with isoflurane (Baxter, 2.5% vaporized in an 80:20 mixture of oxygen and air), and the hair in the neck and scalp was removed with hair removal lotion (Nair). The frontoparietal skull was exposed and the mouse head was immobilized in a custom-made stereotactic holder. A catheter was placed into the tail vein with a 30-gauge needle attached to PE-10 tubing (Becton Dickinson). The imaging system was composed of an LSM510 NLO multiphoton microscope (Carl Zeiss) driven by a Chameleon XR

Ti:Sapphire laser (Coherent) tuned to 880 nm, and an inverted microscope (Axiovert 200; Carl Zeiss) equipped with a ×40 water immersion objective (Achromplan IR, NA 0.8; Carl Zeiss). Fluorescent cells were detected through a bandpass emission filter at 525/50 nm (for EGFP). Vessels were visualized by injecting 70 kDa Texas Red conjugated-dextran (detected using a 620/60 nm filter) intravenously immediately before imaging. Image stacks were collected at 3-μm vertical step size at a depth of 100–150 μm below the skull bone surface. For three-dimensional videos, four sequential image stacks were acquired at 3 μm z spacing to cover a volume of 154 × 154 × 9.0 μm. The time resolution was 1 min. Raw imaging data were processed with Imaris (Bitplane) with a Gaussian filter for noise reduction.

**Ovariectomy and FTY720 treatment.** Twelve 12-week-old female, ovariectomized or sham-operated, C57BL/6J mice were purchased from Charles River Laboratories. Mice were injected intraperitoneally either with FTY720 (3 mg kg<sup>-1</sup>, Cayman Chemical) dissolved in a vehicle (PBS containing 5% acidified DMSO and 30% fatty acid-free BSA) or with vehicle only, daily for 4 weeks. The mice were then killed and femurs were excised and fixed. Uteri of all the animals were excised and weighed to evaluate the effect of ovariectomy.

**Bone histomorphometry.** Trabecular bone morphometry within the metaphyseal region of distal femur was quantified using micro-CT (μCT40, Scanco Medical AG). Volumetric regions for trabecular analysis were selected within the endosteal borders to include the central 80% of vertebral height and secondary spongiosa of femoral metaphyses located ~6% of the total length from the growth plate. Trabecular morphometry was characterized by measuring the bone volume fraction (bone volume/total volume), trabecular thickness and trabecular number. To assess the attachment of osteoclasts to the bone surface, osteoclasts and bone trabeculae were fluorescently visualized in sections by two-photon microscopy. To identify osteoclasts, fluorescence based TRAP staining<sup>20</sup> was performed. Collagen-enriched bone matrices could be visualized using the second harmonic emission from collagen fibres excited by infrared lasers. Sections were examined using a two-photon laser microscope (laser was tuned to 780 nm). Fluorescence was detected through bandpass emission filters at 370–450 nm (for second harmonic emission from bone matrices) or 525/50 nm (for TRAP staining). Image analysis was performed using a fully automated segmentation approach (no manual tracing of cell/tissue boundaries or other user intervention required) and the 'osteoclast attachment ratio' was calculated as the ratio 'bone surface attached by osteoclasts/total bone surface' (F.K., submitted manuscript).

**Flow cytometry.** All reagents were purchased from BD PharMingen. To examine the composition of PMCs, blood was collected from the retro-orbital plexus with a heparinized glass pipette, from mice treated intraperitoneally 4 h previously with FTY720 (3 mg kg<sup>-1</sup>) or vehicle (Fig. 4), or treated intravenously 1 h previously with SEW2871 (5 mg kg<sup>-1</sup>) or vehicle (Supplementary Fig. 4). After removing the red blood cells using ACK lysis buffer (Invitrogen), cells were stained with Alexa647-conjugated anti-F4/80 (Fig. 4), or FITC-conjugated anti-CD11b and PE-Cy7-conjugated anti-CD3 (Supplementary Figs 4 and 9), using conventional methods. Flow cytometric data were collected on a FACSCalibur (Becton Dickinson) and analysed with FlowJo software (TreeStar).

# Chromatin remodelling factor *Mll1* is essential for neurogenesis from postnatal neural stem cells

Daniel A. Lim<sup>1,2,3\*</sup>, Yin-Cheng Huang<sup>1,2\*†</sup>, Tomek Swigut<sup>4</sup>, Anika L. Mirick<sup>5</sup>, Jose Manuel Garcia-Verdugo<sup>6,7</sup>, Joanna Wysocka<sup>4</sup>, Patricia Ernst<sup>5</sup> & Arturo Alvarez-Buylla<sup>1,2</sup>

Epigenetic mechanisms that maintain neurogenesis throughout adult life remain poorly understood<sup>1</sup>. Trithorax group (trxG) and Polycomb group (PcG) gene products are part of an evolutionarily conserved chromatin remodelling system that activate or silence gene expression, respectively<sup>2</sup>. Although PcG member *Bmi1* has been shown to be required for postnatal neural stem cell self-renewal<sup>3,4</sup>, the role of trxG genes remains unknown. Here we show that the trxG member *Mll1* (*mixed-lineage leukaemia 1*) is required for neurogenesis in the mouse postnatal brain. *Mll1*-deficient subventricular zone neural stem cells survive, proliferate and efficiently differentiate into glial lineages; however, neuronal differentiation is severely impaired. In *Mll1*-deficient cells, early proneural *Mash1* (also known as *Ascl1*) and gliogenic *Olig2* expression are preserved, but *Dlx2*, a key downstream regulator of subventricular zone neurogenesis, is not expressed. Overexpression of *Dlx2* can rescue neurogenesis in *Mll1*-deficient cells. Chromatin immunoprecipitation demonstrates that *Dlx2* is a direct target of MLL in subventricular zone cells. In differentiating wild-type subventricular zone cells, *Mash1*, *Olig2* and *Dlx2* loci have high levels of histone 3 trimethylated at lysine 4 (H3K4me3), consistent with their transcription. In contrast, in *Mll1*-deficient subventricular zone cells, chromatin at *Dlx2* is bivalently marked by both H3K4me3 and histone 3 trimethylated at lysine 27 (H3K27me3), and the *Dlx2* gene fails to properly activate. These data support a model in which *Mll1* is required to resolve key silenced bivalent loci in postnatal neural precursors to the actively transcribed state for the induction of neurogenesis, but not for gliogenesis.

Neural stem cells (NSCs) and neurogenesis persist throughout life in the subventricular zone (SVZ) and dentate gyrus of the hippocampus. The *Mll1* histone methyltransferase is expressed in adult SVZ cells<sup>5</sup> as well as embryonic SVZ and ventricular zone (Supplementary Fig. 1). In development, *Mll1* regulates the epigenetic maintenance of homeotic gene expression patterns<sup>2</sup>. Human *MLL* is a proto-oncogene, with chromosomal translocations resulting in MLL fusion proteins that produce human leukaemias with mixed lineage identity<sup>6</sup>. Mouse *Mll1* is required for embryonic haematopoiesis, and most *Mll1*-null mice die between embryonic day (E)10.5 and E12 (ref. 7). We therefore used a conditional knockout 'floxed' allele of *Mll1* (*Mll1*<sup>F/F</sup>)<sup>8</sup>.

To delete *Mll1* from a subset of NSCs, we used the transgenic mouse *hGFAP-Cre*<sup>9</sup>, which exhibits excision of floxed alleles in precursors of the hippocampal dentate gyrus, cerebellar granular cells, and SVZ NSCs at E13.5 (ref. 10). *hGFAP-Cre;Mll1*<sup>F/F</sup> mice were born

at the expected Mendelian ratios and were indistinguishable from their wild-type and *hGFAP-Cre;Mll1*<sup>F/+</sup> littermates (hereafter referred to as controls). However, by postnatal day (P)15, *hGFAP-Cre;Mll1*<sup>F/F</sup> mice developed progressive growth retardation and ataxia, and the mice usually died between P25 and P30. We therefore initially analysed *hGFAP-Cre;Mll1*<sup>F/F</sup> mice and controls at P23–P25.

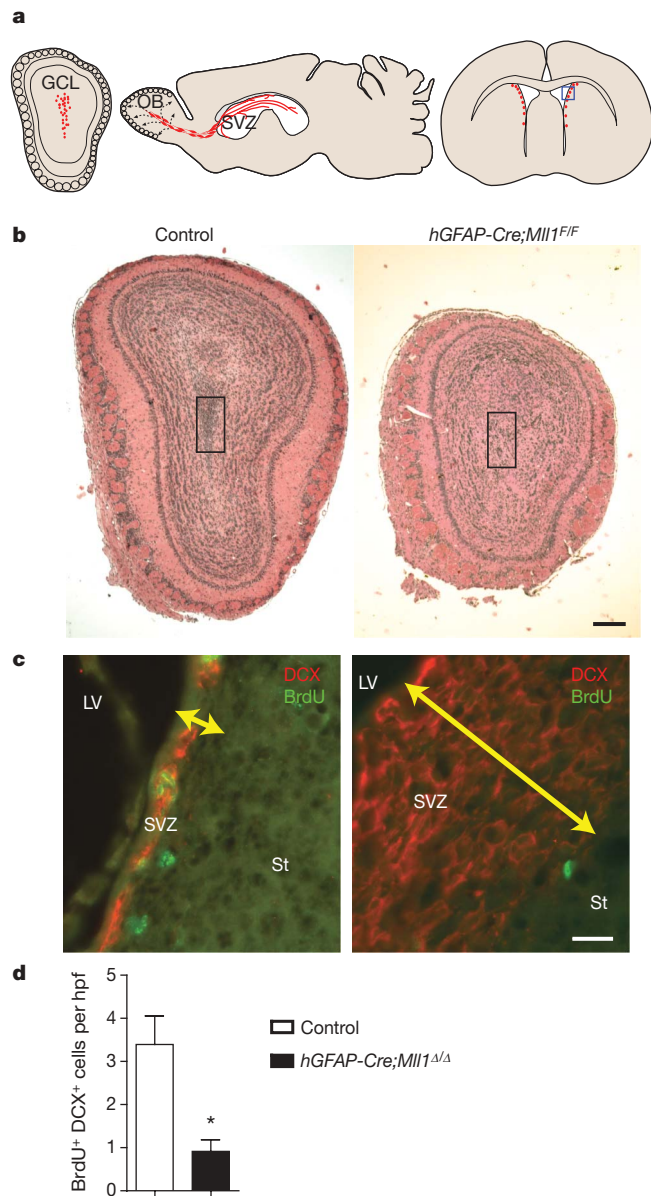
All brain regions that undergo considerable postnatal neurogenesis—including the cerebellar granule cell layer, hippocampal dentate gyrus (Supplementary Fig. 2) and the olfactory bulb (Fig. 1b)—showed a marked reduction in the size and the number of neurons in *hGFAP-Cre;Mll1*<sup>F/F</sup> mice. This suggested a common requirement for *Mll1* in neurogenesis. To investigate the role of *Mll1* in a NSC population, we focused on the SVZ-olfactory-bulb system. Throughout life, NSCs in the SVZ generate large numbers of neuroblasts that migrate to the olfactory bulb<sup>11–13</sup> where they differentiate into interneurons (Fig. 1a, schematic). To evaluate the rate of SVZ neurogenesis, we injected animals with the cell-cycle marker 5-bromo-2-deoxyuridine (BrdU) 1 h before being culled and then co-immunostained sections for BrdU and the neuroblast marker doublecortin (DCX). *hGFAP-Cre;Mll1*<sup>F/F</sup> mice had 3–4-fold fewer BrdU<sup>+</sup>DCX<sup>+</sup> SVZ cells than the controls (Fig. 1c, d). Despite this decreased rate of BrdU incorporation, *hGFAP-Cre;Mll1*<sup>F/F</sup> mice had an expanded SVZ (Fig. 1c, right, yellow double-headed arrow) containing DCX<sup>+</sup> cells; this accumulation developed after P7 (Supplementary Fig. 3) and could be explained if *Mll1*-deficient progenitor cells have impaired and/or abnormal migration, resulting in cell accumulation. Although *Mll1*-deficient progenitors retained expression of neuroblast markers such as DCX and Tuj1 (Supplementary Fig. 3), they also possessed ultrastructural characteristics of intermediate, transit-amplifying cells (Supplementary Fig. 5k, l). Indeed, neuroblast chain migration in *hGFAP-Cre;Mll1*<sup>F/F</sup> mice and from SVZ explants<sup>14</sup> was severely disorganized, resulting in a 40% reduction in the migration distance *in vitro* (Supplementary Fig. 4). Thus, postnatal neuronal addition in the olfactory bulb was abrogated by a decreased rate of neurogenesis as well as by impaired neuroblast migration.

In addition to producing neurons, SVZ NSCs generate astrocytes and oligodendrocytes<sup>15–17</sup>. Glial development in *hGFAP-Cre;Mll1*<sup>F/F</sup> mice was not impaired. In fact, expression of the astrocyte marker GFAP was increased in the SVZ of *hGFAP-Cre;Mll1*<sup>F/F</sup> mice (Fig. 2a) as well as in other brain regions (Supplementary Fig. 5e–h). The SVZ ependymal layer develops postnatally<sup>18</sup>, and these cells formed normally in *hGFAP-Cre;Mll1*<sup>F/F</sup> mice (Fig. 2b and Supplementary Fig. 5i, j). OLIG2 is expressed in developing and mature oligodendrocytes<sup>19,20</sup>, and OLIG2 expression in white matter was similar in

<sup>1</sup>Department of Neurological Surgery, <sup>2</sup>Institute for Regeneration Medicine, and <sup>3</sup>Veteran's Affairs Medical Center, University of California, San Francisco, 505 Parnassus Street M779, San Francisco, California 94143, USA. <sup>4</sup>Department of Chemical and Systems Biology, Department of Developmental Biology Stanford University School of Medicine, Stanford, California 94305, USA. <sup>5</sup>Department of Genetics, Dartmouth Medical School, Hanover, New Hampshire 03755, USA. <sup>6</sup>Laboratorio de Neurobiología Comparada, Instituto Cavanilles, Universidad de Valencia, Valencia 46012, Spain. <sup>7</sup>Laboratorio de Morfología Celular, Centro de Investigación Príncipe Felipe, CIBERNED, Valencia 46012, Spain. <sup>†</sup>Present address: Department of Neurosurgery, Graduate Institute of Clinical Medical Science, ChangGung University, Kwei-Shan, Tao-yuan, Taiwan.

\*These authors contributed equally to this work.



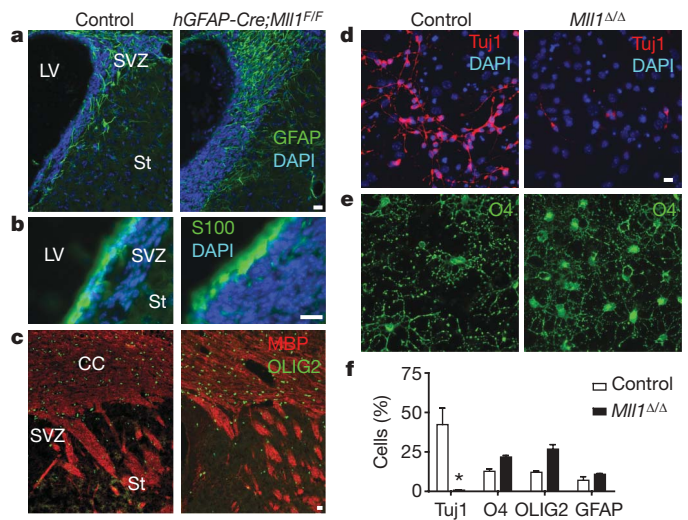


**Figure 1 | *Mll1* is required for normal SVZ-olfactory bulb neurogenesis.**

**a**, Schematics of SVZ-olfactory bulb neurogenesis. Left, coronal section of the olfactory bulb (OB) indicating the region where newly born neuroblasts (red dots) initially arrive from the SVZ. GCL, granule cell layer. Middle, sagittal section showing paths of neuroblast migration from the SVZ to the olfactory bulb. Right, coronal section indicating the germinal SVZ (red dots); the blue box indicates regions shown in **c**. **b**, Haematoxylin and eosin (H&E)-stained coronal sections through the P25 olfactory bulb of control (left) and *hGFAP-Cre;Mll1<sup>F/F</sup>* (right) mice. The black box indicates the olfactory bulb core comprised of recently born neuroblasts. **c**, DCX (red) and BrdU (green) immunohistochemistry of the SVZ is shown. The size of the SVZ is indicated by a yellow double-headed arrow. LV, lateral ventricle; St, striatum. **d**, Quantification of BrdU<sup>+</sup> DCX<sup>+</sup> SVZ cells. hpf, high power field. Error bars, s.e.m.; three mice per group; \**P* = 0.025. Scale bars, 200 μm (**b**) and 20 μm (**c**).

*hGFAP-Cre;Mll1<sup>F/F</sup>* mice and controls (Fig. 2c, green). Furthermore, oligodendrocyte myelination of major axon tracts in *hGFAP-Cre;Mll1<sup>F/F</sup>* mice appeared normal as assessed by FluoroMyelin staining (Supplementary Fig. 5a–d) and myelin basic protein (MBP) immunohistochemistry (Fig. 2c, red). However, it was possible that local, non-SVZ glial progenitors had compensated for a defective SVZ stem cell population.

Therefore, to specifically examine the developmental potential of SVZ stem cells, we used SVZ monolayer NSC cultures<sup>21</sup>. MLL1 was

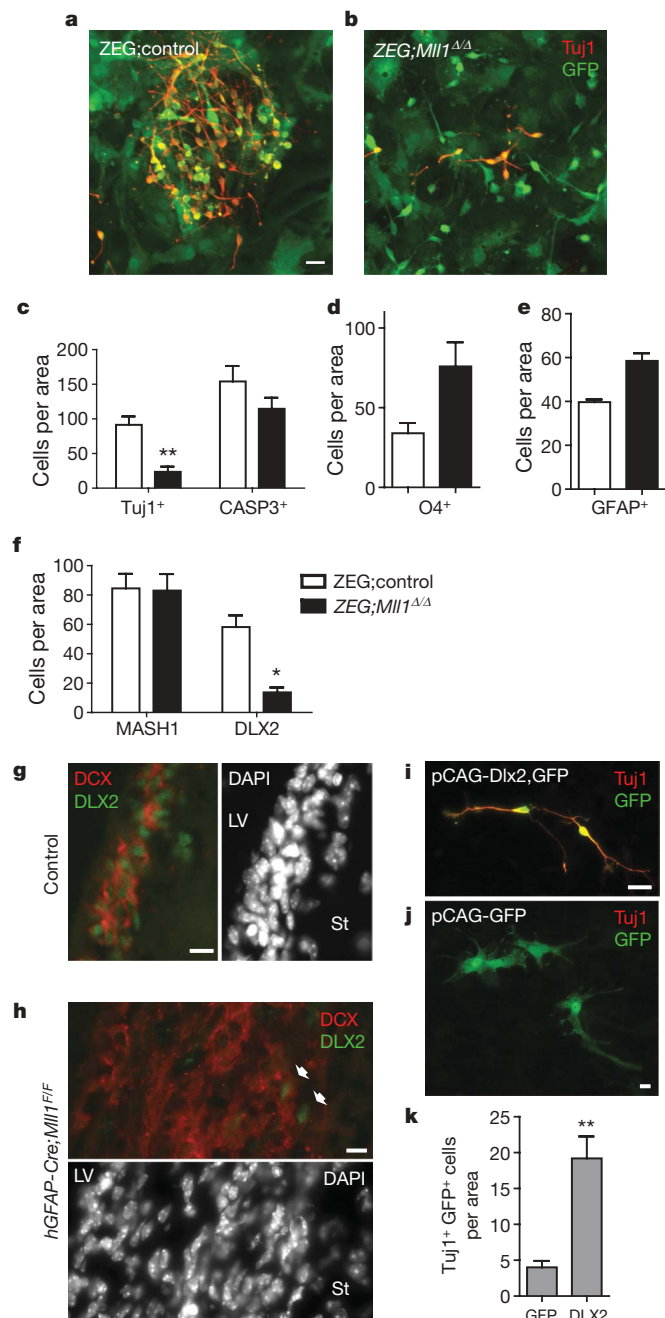


**Figure 2 | *Mll1*-deletion impairs postnatal SVZ-olfactory bulb neurogenesis but not gliogenesis.**

**a**, Immunohistochemistry for the astrocyte marker GFAP (green) in P25 coronal brain sections of control (left) and *hGFAP-Cre;Mll1<sup>F/F</sup>* (right) mice. Nuclei are counterstained with 4,6-diamidino-2-phenylindole (DAPI; blue). **b**, Immunohistochemistry for the S100<sup>+</sup> (green) ependymal cells. **c**, Immunohistochemistry for markers of oligodendrocytes, OLIG2 (green) and MBP (red). CC, corpus callosum. **d**, SVZ cultures after 4 days of differentiation from control (left) and *Mll1<sup>Δ/Δ</sup>* (right) mice immunostained for the neuronal marker TuJ1 (red). **e**, O4<sup>+</sup> (green) oligodendrocytes in the same fields of view as those in **d**. **f**, Quantification of cell differentiation. Error bars, s.e.m. of triplicate cultures; \**P* = 0.016. Scale bars, 20 μm.

expressed in control SVZ cultures, whereas cultures from *hGFAP-Cre;Mll1<sup>F/F</sup>* mice were 95–96% MLL1-deleted (Supplementary Fig. 6a–e). *Mll1<sup>Δ/Δ</sup>* and control SVZ cultures had similar proliferation rates (27.5% ± 0.7% BrdU<sup>+</sup> cells in *Mll1<sup>Δ/Δ</sup>* cultures versus 26.7% ± 2.4% in controls; mean ± s.e.m.). In differentiation conditions, *Mll1<sup>Δ/Δ</sup>* cultures produced >20-fold fewer TuJ1<sup>+</sup> cells compared to controls (Fig. 2d, f). *Mll1<sup>Δ/Δ</sup>* cultures had comparable numbers of pycnotic cells under both proliferative (79% of control, 31% s.d.) and differentiation (68% of control, 15% s.d.) conditions, suggesting that *Mll1* was not required for SVZ cell survival. Similar to our observations *in vivo*, *Mll1<sup>Δ/Δ</sup>* cultures efficiently produced GFAP<sup>+</sup> astrocytes and O4<sup>+</sup> and OLIG2<sup>+</sup> oligodendrocytes (Fig. 2e, f and Supplementary Fig. 6i, j). Thus, *Mll1<sup>Δ/Δ</sup>* SVZ NSC cultures had a similar proliferation rate and comparable cell death rate, but they produced fewer neuronal cells while remaining gliogenic.

Because *Mll1* was deleted in radial glial precursors at E13.5, it was possible that many transcriptional changes had ‘accumulated’ before the genesis of SVZ NSCs. Therefore, we next studied the effect of *Mll1* deletion after postnatal SVZ NSC genesis. To accomplish this, we derived SVZ NSC cultures from P6–P7 *Mll1<sup>F/F</sup>* or *Mll1<sup>+/+</sup>* mice that also carried the ZEG reporter transgene; ZEG expresses green fluorescent protein (GFP) in cells that have undergone Cre-mediated recombination<sup>22</sup>. Cultures were then infected with an adenovirus expressing Cre, and 48–72 h later GFP<sup>+</sup> cells were isolated by fluorescent activated cell sorting (FACS) (Supplementary Fig. 7a). Ninety-five to ninety-eight per cent of GFP<sup>+</sup> *Mll1<sup>F/F</sup>;ZEG* cells (hereafter referred to as ZEG;*Mll1<sup>Δ/Δ</sup>* cells) did not express MLL1; conversely, 100% of GFP<sup>+</sup> *Mll1<sup>+/+</sup>;ZEG* cells (hereafter referred to as ZEG; control cells) expressed MLL1 protein (Supplementary Fig. 7b–e). There was no significant difference in cell death between proliferating ZEG;*Mll1<sup>Δ/Δ</sup>* and ZEG; control stem cell cultures as assessed by propidium iodide and annexin A5 staining (Supplementary Fig. 7f). After FACS isolation, ZEG;*Mll1<sup>Δ/Δ</sup>* and ZEG; control cells also had similar BrdU incorporation rates (Supplementary Fig. 7g). Under differentiation conditions, ZEG;*Mll1<sup>Δ/Δ</sup>* SVZ cells produced three-fold fewer TuJ1<sup>+</sup> cells than ZEG; control cells (Fig. 3a–c).



**Figure 3 | *Mll1*-dependent DLX2 expression is required for postnatal SVZ neurogenesis.** **a**, Immunocytochemistry for Tuj1<sup>+</sup> (red) neuroblasts in FACS-isolated GFP<sup>+</sup> ZEG;control SVZ cells after differentiation. **b**, ZEG;*Mll1*<sup>Δ/Δ</sup> cultures stained for Tuj1. **c**, Quantification of Tuj1<sup>+</sup> neuronal differentiation and activated caspase 3<sup>+</sup> cells after 4 days of differentiation. **d**, **e**, Quantification of glial differentiation. The number of O4<sup>+</sup> oligodendrocytes (**d**) and GFAP<sup>+</sup> astrocytes (**e**) was counted after 4–5 days of differentiation. **f**, Quantification of MASH1<sup>+</sup> and DLX2<sup>+</sup> cells after 2 days of differentiation. **g**, **h**, Immunohistochemistry for DLX2 (green) and DCX (red) in SVZ coronal brain sections of control (**g**) and *hGFAP-Cre;Mll1*<sup>F/F</sup> mice (**h**). **i–k**, Enforced *Dlx2* expression can rescue neurogenesis in *Mll1*<sup>Δ/Δ</sup> SVZ cultures. Immunocytochemistry for Tuj1 (red) and GFP (green) after transfection of both pCAG-Dlx2 and pCAG-GFP plasmids (**i**) and pCAG-GFP plasmid alone (**j**). **k**, Quantification of neuronal lineage rescue by *Dlx2* transfection. **g**, **h**, Scale bars, 10 μm; **a**, **b**, **i**, **j**, **k**, Scale bars, 20 μm.

Differentiating ZEG;*Mll1*<sup>Δ/Δ</sup> cells had a trend towards fewer activated caspase 3<sup>+</sup> cells (Fig. 3c), indicating that cell death did not account for the decreased neurogenesis. Furthermore, ZEG;*Mll1*<sup>Δ/Δ</sup> cells produced glial populations efficiently; in fact, there was approximately

twofold more O4<sup>+</sup> oligodendrocytes and 50% more GFAP<sup>+</sup> astrocytes (Fig. 3d, e and Supplementary Fig. 7h, i) in ZEG;*Mll1*<sup>Δ/Δ</sup> SVZ cultures. Thus, *Mll1* was required after the genesis of SVZ stem cells specifically for the neuronal lineage.

Because *Mll1* is important for the epigenetic regulation of specific gene expression, we next examined the expression of genes important for SVZ neurogenesis. NSC markers SOX2 (ref. 23) and Nestin were present in 100% of cells in both ZEG;*Mll1*<sup>Δ/Δ</sup> and ZEG;control cells under proliferation conditions (Supplementary Fig. 7j, k). During early differentiation, MASH1 expression, a bHLH factor important for SVZ neurogenesis and oligodendroglialogenesis<sup>24</sup>, was also not altered in ZEG;*Mll1*<sup>Δ/Δ</sup> cultures (Fig. 3f and Supplementary Fig. 8a, b). However, DLX2, a homeodomain-containing transcription factor important for olfactory bulb interneuron development and migration in the embryo<sup>25</sup>, was decreased ~fourfold in ZEG;*Mll1*<sup>Δ/Δ</sup> cultures (Fig. 3f and Supplementary Fig. 8c, d). *In vivo* immunostaining showed a similar impairment of DLX2 expression in the SVZ of *hGFAP-Cre;Mll1*<sup>F/F</sup> mice (Fig. 3g, h). Normally, *Dlx2* is expressed in transit-amplifying cells and is then maintained in migratory neuroblasts. *hGFAP-Cre;Mll1*<sup>F/F</sup> mice still had DLX2 in a few cells at the base of the mass of DCX<sup>+</sup> cells, however, most these cells did not express this transcription factor.

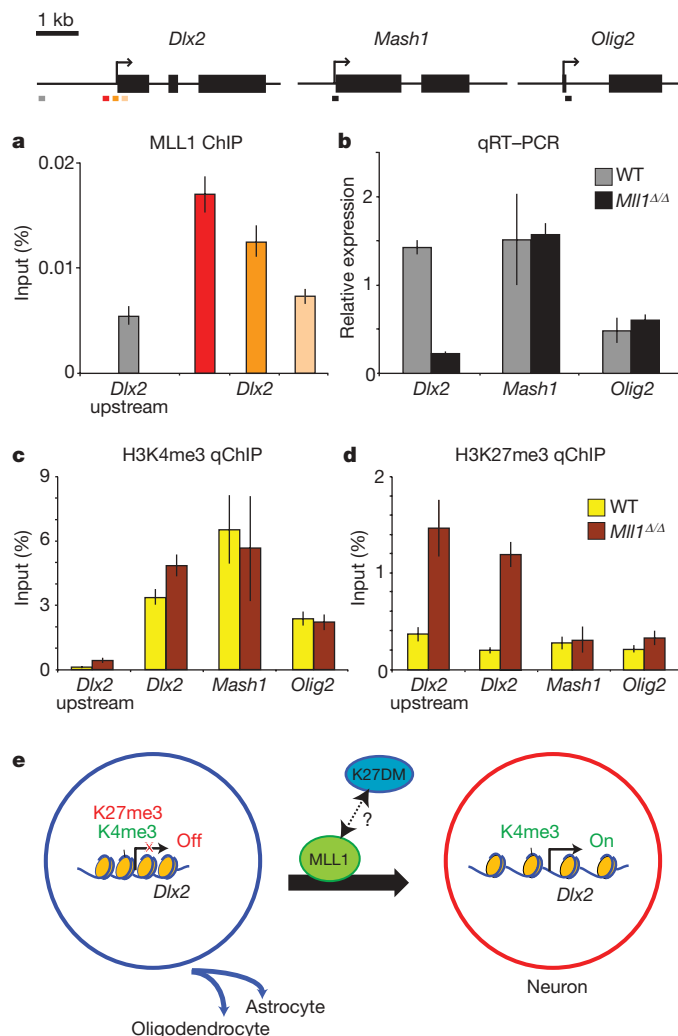
To demonstrate that *Dlx2* is a key developmental regulator for *Mll1*-dependent neurogenesis, we co-transfected *Dlx2* and GFP expression plasmids into *Mll1*<sup>Δ/Δ</sup> SVZ cultures and induced differentiation. Transfection of the GFP plasmid alone was performed as a control. Concordance of GFP and DLX2 expression in co-transfected cultures was ~85% (Supplementary Fig. 9a, b). *Mll1*<sup>Δ/Δ</sup> SVZ cells co-transfected with *Dlx2* and GFP had a fourfold increase in the number of Tuj1<sup>+</sup> GFP<sup>+</sup> cells (Fig. 3i, quantification in panel k) after 4 days of differentiation as compared to the GFP-transfected controls (Fig. 3j).

To determine whether *Dlx2* is a direct target of MLL1, we performed chromatin immunoprecipitation (ChIP) from differentiating SVZ NSC cultures with anti-MLL1 antibodies. We found that MLL1 specifically bound to the *Dlx2* promoter region and the region immediately downstream from the transcriptional start site, but not to the chromatin 1 kb upstream (Fig. 4a).

Methylation of histone lysine residues is a critical determinant of active and silent gene expression states. H3K4me3 correlates strongly with active transcription whereas H3K27me3 is associated with gene silencing. Chromatin regions containing high levels of both H3K4me3 and H3K27me3 have been termed 'bivalent domains' and are silenced but thought to be 'poised' for activation<sup>26</sup>. MLL contains a catalytic SET domain that can methylate histone H3 at K4 (ref. 27). MLL family members can also recruit H3K27-specific histone demethylases<sup>28,29</sup>. Thus MLL proteins possess two non-mutually exclusive mechanisms for promoting transitions between transcriptionally restrictive and permissive chromatin states. We therefore investigated H3 methylation patterns at the promoter regions of *Dlx2*, *Mash1* and *Olig2*. In differentiating SVZ cells, *Dlx2* expression was *Mll1*-dependent whereas *Mash1* and *Olig2* expression were not (Fig. 4b). ChIP analysis of wild-type cells demonstrated that there were high levels of H3K4me3 and low levels of H3K27me3 at all three loci, coherent with their transcriptionally active state (Fig. 4c, d, yellow bars). Surprisingly, loss of MLL1 did not affect H3K4me3 at any of the analysed loci, suggesting that other H3K4 methyltransferases are important for maintenance of H3K4me3 at the *Dlx2* locus. In *Mll1*<sup>Δ/Δ</sup> cells, H3K27me3 levels were strongly increased at *Dlx2*, but not at *Mash1* or *Olig2* loci (Fig. 4d, brown bars). Moreover, in the absence of MLL1, we observed H3K27me3 spreading 1 kb upstream of the transcriptional start site (Fig. 4d), consistent with previous descriptions of bivalent loci<sup>30</sup>. Thus, in the absence of MLL1, the *Dlx2* locus is bivalent in differentiating SVZ NSCs.

Bivalent domains mark developmentally important loci in pluripotent and multipotent cells<sup>26</sup>. Upon differentiation into lineage-specific precursors, many bivalent domains are resolved into H3K4me3 or H3K27me3 monovalent domains. However, some loci





**Figure 4** | *Dlx2* is trimethylated at both H3K4 and H3K27 in *Mll1*<sup>Δ/Δ</sup> cells. At the top is a schematic indicating the location of the primer sets used for qChIP. **a**, MLL1 qChIP of the *Dlx2* locus. **b**, qRT-PCR analysis of *Dlx2*, *Mash1* and *Olig2* in wild-type (grey bars) and *Mll1*<sup>Δ/Δ</sup> (black bars) cells during early differentiation. **c**, qChIP analysis of H3K4me3 levels at *Dlx2*, *Mash1* and *Olig2* loci. **d**, qChIP for H3K27me3 levels. **e**, Model of *Mll1* function in the specification of the neuronal lineage from NSCs. NSCs have bivalent chromatin domains at key neurogenic genes (for example, *Dlx2*). In this state, precursors can form astrocytes and oligodendrocytes (blue arrows). In order for neurogenesis to proceed (red arrow), MLL1 is required for the resolution of specific bivalent loci, possibly by recruiting H3K27-specific demethylases (K27DM). Error bars, s.d.; *n* = 3.

remain bivalent<sup>30</sup>, possibly reflecting remaining gene expression plasticity. The presence of bivalent domains in tissue-specific stem cell populations suggests that there is a requirement for H3K27me3 demethylase activity at specific loci throughout development, and, in the case of SVZ neurogenesis, into adulthood. Our data are consistent with a model in which MLL1 function, perhaps mediated by H3K27me3 demethylase recruitment, is essential for bivalent domain resolution at *Dlx2* during neurogenesis (Fig. 4e). In the absence of MLL1, the *Dlx2* locus remains bivalent and therefore silenced. Deletion of *Mll1* in broader and/or earlier populations of NSCs in development may reveal a similar requirement of *Mll1* for embryonic neurogenesis.

Taken together, our results indicate that for lifelong neurogenesis, *Mll1* is required by neural precursors to make the epigenetic transition to the neuronal lineage by mediating chromatin modifications at specific loci. Future analysis of the direct targets of MLL and bivalent loci in NSCs at different stages of development may lead

to an epigenetic description of NSC lineage potential and a transcriptional program instructive for neurogenesis.

## METHODS SUMMARY

*Mll1*<sup>F/F</sup> mice were maintained and genotyped as described<sup>8</sup>. For immunocytochemistry, frozen sections were used. Tissue preparation for electron microscopy was performed with standard techniques<sup>13</sup>. For cell culture, mouse SVZ monolayer cultures were derived and grown essentially as previously described<sup>21</sup>. FACS was performed with a FACSARIA cell sorter (BD Biosciences). Annexin A5 staining was quantified with a FACS Vantage flow cytometer (BD Biosciences). Transfection of SVZ cells was performed with Lipofectamine (Invitrogen). Paraformaldehyde-fixed cultures were immunostained using standard protocols. For quantification of cell cultures, at least five non-overlapping high-power fields of view were analysed. SVZ cells *in vivo* were quantified from multiple non-overlapping confocal optical sections obtained with a Zeiss Axiovert 200M. Statistical analysis was performed using GraphPad Prism software; unpaired *t*-tests were used for the two-way comparisons. For quantitative PCR with reverse transcription (qRT-PCR), RNA was isolated with Trizol (Invitrogen), treated with DNase, and reverse transcribed using VILIO Superscript (Invitrogen). qPCR was performed with the Roche LC480 using SybrGreen (Roche). Relative gene expression was normalized to six housekeeping genes. For quantitative ChIP (qChIP), chromatin was prepared from cells fixed with 1% formaldehyde and then sheared by sonication. Chromatin was incubated overnight with the indicated antibodies, and then collected by incubation with Protein A Dynabeads (Invitrogen). DNA eluted from the washed immune complexes was extracted, precipitated and then subjected to qPCR analysis with SybrGreen. Recovery of genomic DNA as a percentage input was calculated as the ratio of copy numbers in the precipitated immune complexes to the input control.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 23 July 2007; accepted 15 December 2008.

Published online 11 February 2009.

- Hsieh, J. & Gage, F. H. Epigenetic control of neural stem cell fate. *Curr. Opin. Genet. Dev.* **14**, 461–469 (2004).
- Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B. & Cavalli, G. Genome regulation by polycomb and trithorax proteins. *Cell* **128**, 735–745 (2007).
- Molofsky, A. V. et al. Bmi-1 dependence distinguishes neural stem cell self-renewal from progenitor proliferation. *Nature* **425**, 962–967 (2003).
- Fasano, C. A. et al. shRNA knockdown of *Bmi-1* reveals a critical role for p21-Rb pathway in NSC self-renewal during development. *Cell Stem Cell* **1**, 87–99 (2007).
- Lim, D. A. et al. *In vivo* transcriptional profile analysis reveals RNA splicing and chromatin remodeling as prominent processes for adult neurogenesis. *Mol. Cell. Neurosci.* **31**, 131–148 (2006).
- Daser, A. & Rabbitts, T. H. Extending the repertoire of the mixed-lineage leukemia gene *MLL* in leukemogenesis. *Genes Dev.* **18**, 965–974 (2004).
- Yu, B. D., Hess, J. L., Horning, S. E., Brown, G. A. & Korsmeyer, S. J. Altered Hox expression and segmental identity in *Mll*-mutant mice. *Nature* **378**, 505–508 (1995).
- Jude, C. D. et al. Unique and independent roles for MLL in adult hematopoietic stem cells and progenitors. *Cell Stem Cell* **1**, 324–337 (2007).
- Zhuo, L. et al. hGFAP-cre transgenic mice for manipulation of glial and neuronal function *in vivo*. *Genesis* **31**, 85–94 (2001).
- Han, Y. G. et al. Hedgehog signaling and primary cilia are required for the formation of adult neural stem cells. *Nature Neurosci.* **11**, 277–284 (2008).
- Luskin, M. B. Restricted proliferation and migration of postnatally generated neurons derived from the forebrain subventricular zone. *Neuron* **11**, 173–189 (1993).
- Lois, C. & Alvarez-Buylla, A. Long-distance neuronal migration in the adult mammalian brain. *Science* **264**, 1145–1148 (1994).
- Doetsch, F., Caille, I., Lim, D. A., Garcia-Verdugo, J. M. & Alvarez-Buylla, A. Subventricular zone astrocytes are neural stem cells in the adult mammalian brain. *Cell* **97**, 703–716 (1999).
- Wichterle, H., Garcia-Verdugo, J. M. & Alvarez-Buylla, A. Direct evidence for homotypic, glia-independent neuronal migration. *Neuron* **18**, 779–791 (1997).
- Marshall, C. A., Suzuki, S. O. & Goldman, J. E. Gliogenic and neurogenic progenitors of the subventricular zone: who are they, where did they come from, and where are they going? *Glia* **43**, 52–61 (2003).
- Picard-Riera, N. et al. Experimental autoimmune encephalomyelitis mobilizes neural progenitors from the subventricular zone to undergo oligodendrogenesis in adult mice. *Proc. Natl Acad. Sci. USA* **99**, 13211–13216 (2002).
- Menn, B. et al. Origin of oligodendrocytes in the subventricular zone of the adult brain. *J. Neurosci.* **26**, 7907–7918 (2006).
- Spassky, N. et al. Adult ependymal cells are postmitotic and are derived from radial glial cells during embryogenesis. *J. Neurosci.* **25**, 10–18 (2005).



19. Lu, Q. R. *et al.* Common developmental requirement for Olig function indicates a motor neuron/oligodendrocyte connection. *Cell* **109**, 75–86 (2002).
20. Zhou, Q. & Anderson, D. J. The bHLH transcription factors OLIG2 and OLIG1 couple neuronal and glial subtype specification. *Cell* **109**, 61–73 (2002).
21. Scheffler, B. *et al.* Phenotypic and functional characterization of adult brain neurogenesis. *Proc. Natl Acad. Sci. USA* **102**, 9353–9358 (2005).
22. Novak, A., Guo, C., Yang, W., Nagy, A. & Lobe, C. G. Z/EG, a double reporter mouse line that expresses enhanced green fluorescent protein upon Cre-mediated excision. *Genesis* **28**, 147–155 (2000).
23. Zappone, M. V. *et al.* Sox2 regulatory sequences direct expression of a  $\beta$ -geo transgene to telencephalic neural stem cells and precursors of the mouse embryo, revealing regionalization of gene expression in CNS stem cells. *Development* **127**, 2367–2382 (2000).
24. Parras, C. M. *et al.* *Mash1* specifies neurons and oligodendrocytes in the postnatal brain. *EMBO J.* **23**, 4495–4505 (2004).
25. Long, J. E. *et al.* Dlx-dependent and -independent regulation of olfactory bulb interneuron differentiation. *J. Neurosci.* **27**, 3230–3243 (2007).
26. Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–326 (2006).
27. Milne, T. A. *et al.* MLL targets SET domain methyltransferase activity to *Hox* gene promoters. *Mol. Cell* **10**, 1107–1117 (2002).
28. Lee, M. G. *et al.* Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. *Science* **318**, 447–450 (2007).
29. Swigut, T. & Wysocka, J. H3K27 demethylases, at long last. *Cell* **131**, 29–32 (2007).
30. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank J. Rubenstein for anti-DLX2 antibodies and the pCAG-Dlx2 plasmid, D. Rowitch for anti-OLIG2 antibodies, and Y. Dou and R. Roeder for anti-MLL1 antibodies. This work was supported by the Neurosurgery Research and Education Foundation/American Association of Neurological Surgeons, Sandler Family Foundation, Northern California Institute for Research and Education, and the Clinical and Translational Research Institute at the University of California, San Francisco (D.A.L.), California Institute for Regenerative Medicine New Faculty Award and The Chicago Community Trust Searle Scholar Award (J.W.), and the Goldhirsch Foundation, J.G. Bowes Research Fund, and National Institutes of Health (NIH) 5R37-NS028478 (A.A.-B.).

**Author Contributions** D.A.L. conceived the project, designed and performed experiments, coordinated collaborations, and wrote the manuscript. Y.-C.H. worked on most experiments, quantified all *in vivo* data, and helped prepare the figures. T.S. and J.W. performed ChIP experiments, helped analyse data and contributed ideas. A.L.M. and P.A.E. provided the *MLL1<sup>+/F</sup>* mouse, helped perform preliminary experiments in *MLL1<sup>+/F</sup>* mice and contributed ideas. J.M.G.V. provided electron microscopy data and histological interpretation. A.A.-B. contributed ideas, interpreted results and helped write the manuscript. All authors discussed the results and edited the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to D.A.L. ([limd@neurosurg.ucsf.edu](mailto:limd@neurosurg.ucsf.edu)) or A.A.-B. ([abuylla@stemcell.ucsf.edu](mailto:abuylla@stemcell.ucsf.edu)).

## METHODS

**Animals.** *Mlll*<sup>F/F</sup> mice were maintained and genotyped as described<sup>8</sup>. *hGFAP-Cre*<sup>9</sup>, *ZEG*<sup>22</sup> and *R26R* mice were obtained from Jackson Laboratories. To label proliferating cells, BrdU (10 mg ml<sup>-1</sup> solution, 50 mg kg<sup>-1</sup> body weight, Sigma) was injected intraperitoneally. Experiments were performed in accordance to protocols approved by Institutional Animal Care and Use Committee at UCSF.

**Tissue preparation, immunohistochemistry and *in situ* hybridization.** Brains were fixed by intracardiac perfusion as previously described<sup>13</sup>. Five-micrometre paraffin sections were used for H&E staining. For immunocytochemistry, 12- $\mu$ m frozen sections were used with the following primary antibodies: mouse anti-GFAP (1:500, Chemicon), guinea pig anti-doublecortin (1:500, Chemicon), rat anti-BrdU (1:100, Abcam), anti-MASH (1:100 Pharmingen), rabbit anti-OLIG2 (1:10,000, gift from D. Rowitch), rabbit anti-DLX2 (1:1,000, gift from J. Rubenstein), rabbit anti-activated caspase 3 (1:500, Cell Signal Technology), anti-Nestin (1:500, Chemicon), rabbit anti-SOX2 (Abcam), mouse anti-S100 (1:500, Dako), mouse anti-MLL, N-terminal (1:500, Upstate), mouse anti- $\beta$ III-tubulin (Tuj1 clone, 1:250, Covance), mouse anti-myelin basic protein (1:2,000, Covance), mouse anti-O4 (1:50, Chemicon) and chicken anti-GFP (1:500, Aves). FluoroMyelin (Invitrogen) was used according to the manufacturer's protocols. Sections were blocked with 10% goat serum plus 0.3% Triton X-100 (Sigma) in PBS for 30 min at 25 °C, before primary antibody incubation at 4 °C overnight. BrdU staining was performed as in ref. 13. Goat Alexa-Fluor secondary antibodies (Invitrogen) were used, and nuclei were counterstained with DAPI (Sigma). *Mlll* *in situ* hybridization was as previously described<sup>5</sup>. X-Gal staining and electron microscopy was performed as previously described<sup>13</sup>.

**Cell culture.** Mouse SVZ monolayer cultures were produced essentially as previously described<sup>21</sup>. In brief, SVZ microdissections were dissociated with 0.25% trypsin and trituration. Cells were plated at  $\sim 30,000$  cells cm<sup>-2</sup> in 6-well plates (Corning) in proliferation medium (DMEM/F12/N2, 5% FCS, 20 ng ml<sup>-1</sup> EGF, 20 ng ml<sup>-1</sup> bFGF, 35  $\mu$ g ml<sup>-1</sup> bovine pituitary extract (media and N2 are from Invitrogen; growth factors are from Peprotech; FCS is from Hyclone). Non-attached cells were collected after 1 day and replated into 6-well plates. After  $\sim 7$  days, the cells were confluent and these were routinely passaged 1:2 with 0.25% trypsin. Cells were passaged 4–6 times before use in experiments. Differentiation of cultures was induced by removing the EGF, FGF and FCS from the media<sup>21</sup>. For FACS, cells were dissociated with 0.25% trypsin and passed through a 40- $\mu$ m mesh. A FACSaria (BD Biosciences) cell sorter with a 70- $\mu$ m nozzle was used at the low pressure setting. Cells were collected into DMEM/F12 with 20% FCS. FACS isolated cells were centrifuged (500g, 15 min), resuspended in proliferation medium, and plated at  $\sim 100,000$  cells cm<sup>-2</sup>. For cell proliferation analysis, BrdU was used at 10  $\mu$ M. For immunostaining, cultures were fixed with 4% paraformaldehyde. Primary and secondary antibodies were used as indicated above, with the exception that O4 staining was performed without Triton X-100. For plasmid transfections, SVZ cells were plated at  $\sim 75,000$  cells cm<sup>-2</sup> into 8- or 16-well Lab-Tek CCR2 chamber slides (Nunc) in proliferation medium. The next morning, 5  $\mu$ g cm<sup>-2</sup> of pCAG-GFP or both pCAG-GFP and pCAG-Dlx2 was transfected with Lipofectamine (Invitrogen). Twenty-four hours after transfection, differentiation was induced. Annexin A5

staining with APC-conjugated antibodies (BD Biosciences) was performed as per the manufacturer protocols and quantified on a fluorescent flow cytometer.

**Microscopy and quantification.** For quantification of cell cultures, at least five non-overlapping fields of view were analysed at the fluorescent microscope (Olympus AX70) with  $\times 20$  to  $\times 60$  objectives; in some cases, digital images were captured and immunostained cells were counted using Photoshop (Adobe Systems) or ImageJ (NIH) software. For *in vivo* SVZ cell quantification, we collected  $>3$  non-overlapping confocal images from each tissue section using a Zeiss Axiovert 200M with a  $\times 63$ -oil objective; from each animal, at least three separate tissue sections were analysed.

**qRT-PCR.** Total cellular RNA was isolated by Trizol method (Invitrogen) and quantified using the NanoDrop spectrophotometer. One microgram of total RNA was treated with Turbo DNase (Ambion) then reverse transcribed with VILO Superscript (Invitrogen). Complementary DNA corresponding to 5 ng of total RNA was used as a template in qPCR analysis performed on a Roche LC480 with SybrGreen (Roche). Relative expression for the studied genes was normalized to the mean signals of six mouse housekeeping genes: *Atpaf2*, *Dhps*, *Gapdh*, *Nosip*, *Pdha1* and *Tufm*.

**Quantitative chromatin immunoprecipitation.** qChIP was performed as essentially as previously described<sup>31</sup>. In brief,  $10^7$  cells were fixed for 10–20 min in 1% formaldehyde and then washed with 125 mM glycine. Cells were pelleted by centrifugation at 1,000g. After freeze-thawing, cells were suspended in 1 ml swelling buffer and separated by 20 strokes in a Dounce homogenizer. The cell suspension was pelleted and resuspended in 250  $\mu$ l of lysis buffer, then sonicated in a Diagenode Bioruptor with ten 30 s pulses over a 15 min period at the high-energy setting. The cell lysate was pelleted by centrifugation at  $1.6 \times 10^4$ g for 15 min at 4 °C. The supernatant was diluted tenfold in ice-cold dilution buffer containing protease inhibitors (Roche) and subsequently cleared by centrifugation at  $1.6 \times 10^4$ g for 15 min at 4 °C. An aliquot of supernatant was reserved as the input control. The remaining portion was incubated overnight with the appropriate antibodies (H3K4 tri-methyl, Abcam; H3K27 tri-methyl, Upstate; MLL rabbit polyclonal, gift from D. Allis). Immune complexes were collected by 30 min incubation with Protein A Dynabeads, washed once with dilution buffer, five times with RIPA buffer and once with TE buffer. Complexes were eluted with 1% SDS and 0.84% NaHCO<sub>3</sub> solution. NaCl was added to 150 mM, and the eluates were decrosslinked by incubation at 65 °C overnight. The eluates were treated with RNase and proteinase K, extracted with phenol/chloroform and ethanol precipitated with glycogen as a carrier. An input control was processed in parallel. DNA was dissolved in ultrapure water and subjected to qPCR analysis with the Roche LC480 and SybrGreen. Serial dilutions of mouse genomic DNA was used for standardization. For qChIP and qRT-PCR, error estimates are standard deviations and were propagated by the least square formula. Recovery of genomic DNA as the percentage input was calculated as the ratio of copy numbers in the immunoprecipitate to the input control. Primer sequences are available on request.

31. Wysocka, J. *et al.* WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell* 121, 859–872 (2005).

## LETTERS

# Nicotine binding to brain receptors requires a strong cation- $\pi$ interaction

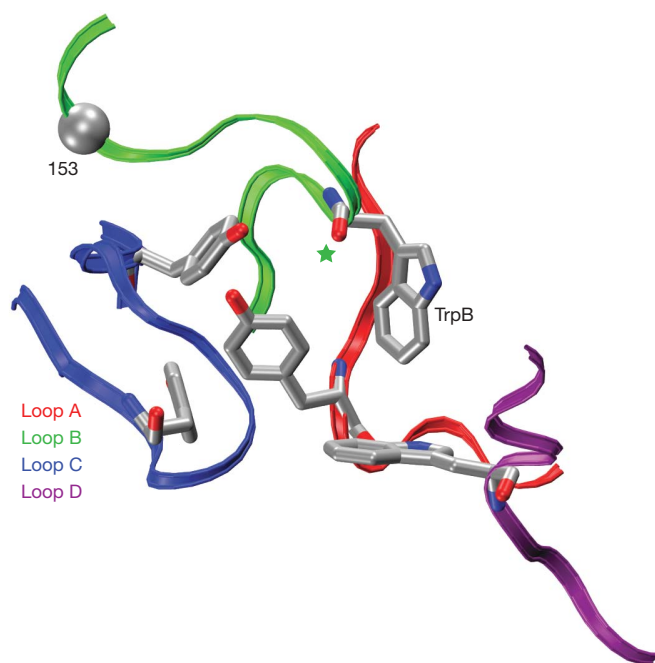
Xinan Xiu<sup>1\*</sup>, Nyssa L. Puskar<sup>1\*</sup>, Jai A. P. Shanata<sup>1</sup>, Henry A. Lester<sup>2</sup> & Dennis A. Dougherty<sup>1</sup>

Nicotine addiction begins with high-affinity binding of nicotine to acetylcholine (ACh) receptors in the brain. The end result is over 4,000,000 smoking-related deaths annually worldwide and the largest source of preventable mortality in developed countries. Stress reduction, pleasure, improved cognition and other central nervous system effects are strongly associated with smoking. However, if nicotine activated ACh receptors found in muscle as potently as it does brain ACh receptors, smoking would cause intolerable and perhaps fatal muscle contractions. Despite extensive pharmacological, functional and structural studies of ACh receptors, the basis for the differential action of nicotine on brain compared with muscle ACh receptors has not been determined. Here we show that at the  $\alpha 4\beta 2$  brain receptors thought to underlie nicotine addiction, the high affinity for nicotine is the result of a strong cation- $\pi$  interaction to a specific aromatic amino acid of the receptor, TrpB. In contrast, the low affinity for nicotine at the muscle-type ACh receptor is largely due to the fact that this key interaction is absent, even though the immediate binding site residues, including the key amino acid TrpB, are identical in the brain and muscle receptors. At the same time a hydrogen bond from nicotine to the backbone carbonyl of TrpB is enhanced in the neuronal receptor relative to the muscle type. A point mutation near TrpB that differentiates  $\alpha 4\beta 2$  and muscle-type receptors seems to influence the shape of the binding site, allowing nicotine to interact more strongly with TrpB in the neuronal receptor. ACh receptors are established therapeutic targets for Alzheimer's disease, schizophrenia, Parkinson's disease, smoking cessation, pain, attention-deficit hyperactivity disorder, epilepsy, autism and depression<sup>1</sup>. Along with solving a chemical mystery in nicotine addiction, our results provide guidance for efforts to develop drugs that target specific types of nicotinic receptors.

Nicotinic acetylcholine receptors (nAChRs) comprise a family of  $\geq 20$  homologous subtypes that mediate fast synaptic transmission throughout the central and peripheral nervous systems<sup>2</sup>. The neuronal receptors are found in the central nervous system (CNS) and autonomic ganglia. Of these, the subtype most strongly associated with nicotine addiction and the target of recently developed smoking cessation drugs is termed  $\alpha 4\beta 2$  (refs 3–7). The high nicotine affinity of  $\alpha 4\beta 2$  receptors, when combined with the ability of nicotine to cross the blood-brain barrier and its favourable pharmacokinetics, allows nicotine at the submicromolar concentrations in tobacco smoke to activate acutely these receptors, providing reward, cognitive sensitization and perhaps other effects. In addition, the high-affinity interaction allows smoked nicotine to act as an intracellular pharmacological chaperone of  $\alpha 4\beta 2$  receptors, leading to the upregulation of receptors thought to underlie effects of chronic exposure<sup>6,8</sup>.

In previous studies of the nAChR of the neuromuscular junction (muscle type), we showed that an important contributor to ACh

binding is a cation- $\pi$  interaction to a specific tryptophan (called TrpB, residue 149, Fig. 1)<sup>9</sup>. These results were subsequently supported by the important series of crystal structures of ACh binding proteins (AChBP)<sup>10,11</sup>. These structures revealed the 'aromatic box' structural motif of Fig. 1, and the aligning residues are predominantly aromatic throughout the Cys-loop family of neurotransmitter-gated ion channels. In other Cys-loop receptors, a cation- $\pi$  interaction between the natural agonist and one of the aromatics is always seen, although its precise location varies<sup>12</sup>. Interestingly, when nicotine activates the muscle-type nAChR, there is no cation- $\pi$  interaction<sup>13</sup>, consistent with its relatively low affinity for this receptor. This suggested that a cation- $\pi$  interaction could discriminate between high-affinity neuronal receptors and low-affinity muscle-type receptors. However, subtle effects must be involved, as the nAChRs of the CNS and neuromuscular junction are homologous throughout most regions of sequence and are essentially identical in the immediate vicinity of the agonist binding site (Supplementary Fig. 1).



**Figure 1 | The binding site of AChBP, thought to resemble that of nAChRs.** Shown are the four principal 'loops' that define the binding site<sup>2</sup>. Also highlighted are TrpB (149) studied here; its backbone carbonyl (green star); and the  $\alpha$  carbon on position 153, which has also been mutated here. Note that loop C contributes two aromatic residues; the other loops each contribute one. The image is of Protein Data Bank file 1I9B (ref. 10).

<sup>1</sup>Divisions of Chemistry and Chemical Engineering and <sup>2</sup>Biology, California Institute of Technology, 1200 East California Boulevard, Pasadena, California 91125, USA.

\*These authors contributed equally to this work.



Here we describe studies of the  $\alpha 4\beta 2$  neuronal receptor. We find a remarkable alteration of binding behaviour: both ACh and nicotine make a strong cation– $\pi$  interaction to TrpB. In addition, a hydrogen bond from nicotine to the backbone carbonyl of TrpB that is weak in the muscle-type is much stronger in the  $\alpha 4\beta 2$  receptor. Taken together, these two noncovalent interactions fully rationalize the differential affinity of nicotine in the brain versus the neuromuscular junction.

A cation– $\pi$  interaction between a drug and a receptor can be revealed by incorporation of a series of fluorinated amino acid analogues (Fig. 2); a consistent trend in receptor response indicates a binding interaction. Such an experiment is enabled by the nonsense suppression methodology for incorporation of unnatural amino acids into receptors and channels expressed in *Xenopus* oocytes. Although we have found the nonsense suppression methodology to be broadly applicable<sup>14,15</sup>, implementing the methodology for study of the  $\alpha 4\beta 2$  neuronal nAChRs proved to be especially challenging, requiring new strategies. The  $\alpha 4\beta 2$  receptors are expressed in *Xenopus* oocytes at inadequately low levels for nonsense suppression experiments. However, recent studies showed that the Leu9'Ala (L9'A) mutation in the M2 transmembrane helix of the  $\alpha 4$  subunit greatly improves expression without altering the pharmacological selectivity of the receptor<sup>16</sup>. (In Cys-loop receptors, the highly homologous M2 sequences are often compared by numbering from the cytoplasmic end, termed position 1'.) Therefore, all studies of  $\alpha 4\beta 2$  described here included this mutation. As with other mutations of L9', the L9'A mutation lowers the agonist concentration for half-maximum response ( $EC_{50}$ ) by influencing receptor gating in ways that are fairly well understood and that do not distort the present analysis of the binding site (some 60 Å from the 9' position)<sup>17,18</sup>. In addition, previous studies of the muscle-type receptor used a comparable mutation at L9', and control experiments established that it did not alter binding trends<sup>9,19</sup>.

The nAChRs are pentameric. The muscle-type receptor has a precise stoichiometry of  $(\alpha 1)_2\beta 1\gamma\delta$ . However, the  $\alpha 4\beta 2$  receptor can have variable stoichiometry. In particular, there are two forms of  $\alpha 4\beta 2$ ,  $(\alpha 4)_2(\beta 2)_3$  and  $(\alpha 4)_3(\beta 2)_2$ , which we refer to hereafter as A2B3 and A3B2, respectively<sup>8,20,21</sup>. Agonist binding sites are at the appropriate  $\alpha$ – $\beta$  interfaces. The A2B3 form has higher sensitivity for nicotine and may be upregulated during chronic exposure to nicotine; our studies have focused on it. Controlling the ratios of messenger RNAs injected into the oocyte can reliably control subunit

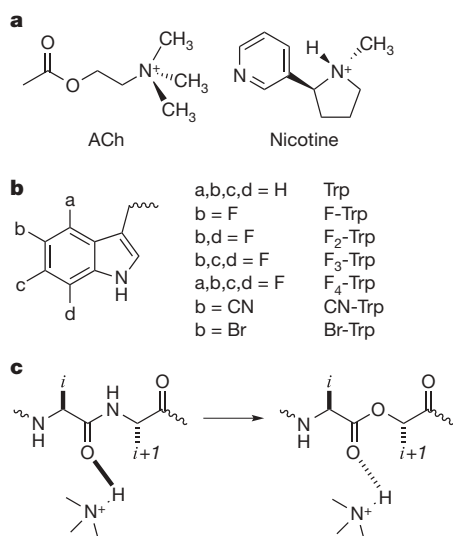
stoichiometry in the wild-type receptor. However, in a nonsense suppression experiment, the subunit that contains the stop codon where the unnatural amino acid has been incorporated can show low and variable expression levels. Therefore we sought a second, independent indicator of the stoichiometry of the  $\alpha 4\beta 2$  receptor. We now report that the A2B3 and A3B2 forms of the  $\alpha 4$ (L9'A) $\beta 2$  receptor show markedly different rectification behaviours. As indicated by either voltage ramp or voltage jump experiments, A2B3 is substantially more inward rectifying than A3B2 (Supplementary Fig. 2). Thus, in all our experiments with unnatural amino acids, the stoichiometries of mutant receptors are monitored by measuring current–voltage relations with voltage jumps. For each mutant receptor studied, we determined the fraction (outward current at +70 mV/inward current at –110 mV), and a value  $\leq 0.1$  establishes the desired A2B3 stoichiometry (Supplementary Table 1 and Supplementary Discussion). With these methodological developments in hand, incorporation of unnatural amino acids into the  $\alpha 4\beta 2$  receptor becomes feasible (Fig. 3).

As shown in Supplementary Table 1 and Fig. 4, a compelling 'fluorination' trend is seen for both ACh and nicotine at TrpB of the  $\alpha 4\beta 2$  receptor. This is in contrast to the results at the muscle-type receptor, in which no such trend is seen for nicotine activation. Further support for an important cation– $\pi$  interaction for both agonists is provided by the large perturbation induced by a cyano (CN) group—which is strongly deactivating in a cation– $\pi$  interaction—compared to a bromo (Br) group, which is roughly isosteric to a cyano group but much less deactivating.

We have also evaluated other residues that constitute the aromatic box of the ACh binding site (Supplementary Table 1). The results for  $\alpha 4\beta 2$  very much parallel our previous findings for the muscle-type receptor (Supplementary Discussion). This indicates that it is specifically the interaction with TrpB that discriminates the two receptor subtypes.

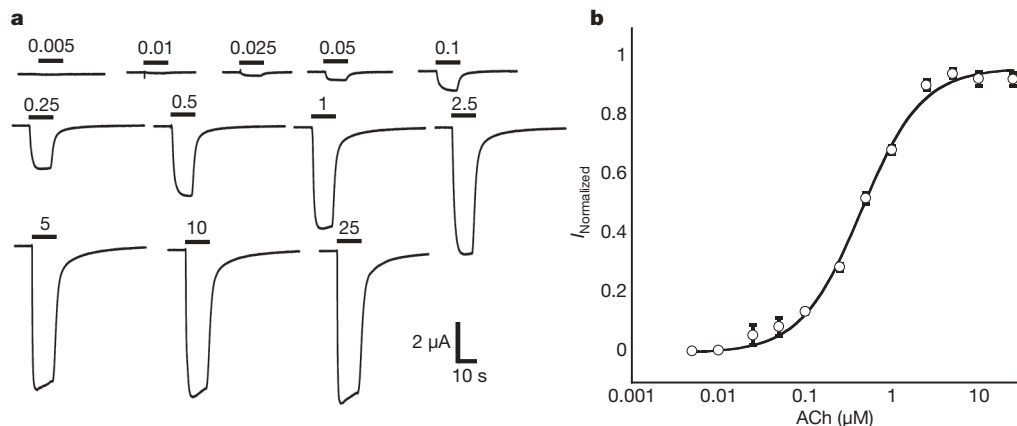
The  $EC_{50}$  values reported here represent a measure of receptor function; shifts in  $EC_{50}$  can result from changes in ligand binding and/or receptor gating properties. By ascribing the results to attenuation of a cation– $\pi$  interaction, we are effectively concluding that it is ligand binding that is being modulated by fluorination, but that conclusion is not incontrovertible. To resolve this ambiguity, we evaluated the gating behaviours of key receptors using single-channel recording. For the wild type and the receptor with 5,6,7-trifluorotryptophan (F<sub>3</sub>-Trp) at TrpB, we compared the probabilities that the channel is open ( $P_{open}$ ) at nicotine concentrations that evoke half-maximal macroscopic steady-state currents ( $EC_{50} = 0.08 \mu\text{M}$  and  $1.2 \mu\text{M}$ , respectively). Any differences between the two  $P_{open}$  values must result from differences in gating behaviours. As suggested by Fig. 5 and as confirmed by further single-channel analysis (Supplementary Fig. 3 and Supplementary Discussion), the wild-type and mutant receptors have  $P_{open}$  values that are essentially indistinguishable. Thus, the shift in  $EC_{50}$  for F<sub>3</sub>-Trp is primarily, if not exclusively, a consequence of changes in binding. Fluorination of TrpB of the  $\alpha 4\beta 2$  (A2B3) receptor primarily has an impact on the sensitivity to nicotine by decreasing nicotine's cation– $\pi$  interaction with this residue.

These results indicate that nicotine is positioned more closely to TrpB in the  $\alpha 4\beta 2$  agonist binding site than in the muscle type. This suggested that another nicotine-binding interaction could also be altered. An important chemical distinction between ACh and nicotine is that only the latter can act as a hydrogen bond donor, through the pyrrolidine  $N^+$ -H (Fig. 2a). Examination of the AChBP crystal structures (Fig. 1)<sup>22</sup> suggested that the backbone carbonyl associated with TrpB could act as the hydrogen bond acceptor, and several groups have shown the importance of this interaction<sup>22–24</sup>. Previously, we probed this potential hydrogen bond in the muscle-type receptor by replacing the ( $i+1$ ) residue with its  $\alpha$ -hydroxy analogue (Fig. 2c). This converts the backbone amide to a backbone ester, which is well established to be a substantially poorer hydrogen bond acceptor. In the muscle-type receptor, this change raised the



**Figure 2 | Agonists and unnatural amino acids considered here.**

**a**, Structures of ACh and nicotine. **b**, Unnatural amino acids considered here. If not indicated, an a, b, c, or d group is H. Br, bromo group; CN, cyano group. **c**, The backbone ester strategy for modulating a hydrogen bond.



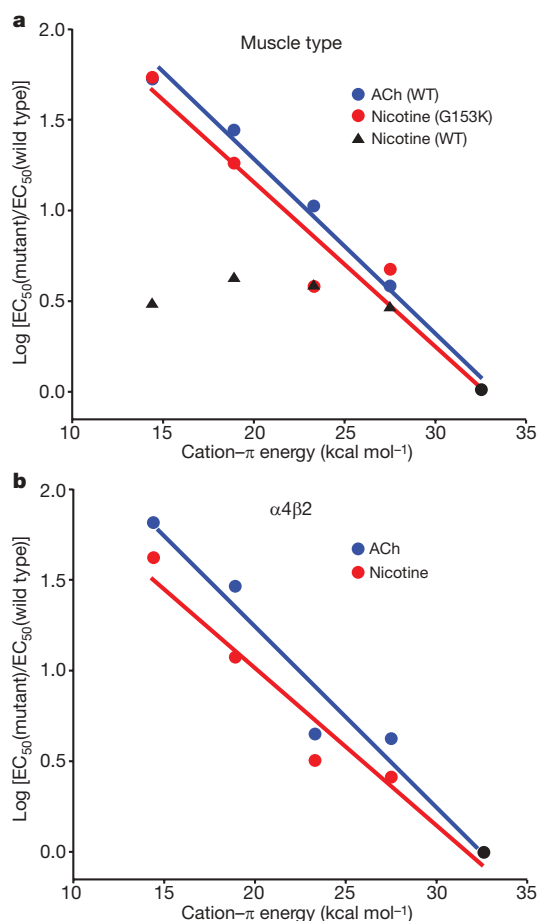
**Figure 3 | Nonsense suppression in the  $\alpha 4\beta 2$  receptor.** Shown is a wild-type recovery experiment, in which Trp is incorporated at the TrpB position. **a**, Representative traces of voltage-clamp currents. Bars represent

application of ACh at concentrations noted. **b**, Fit of data in **a** to the Hill equation. Error bars indicate s.e.m.;  $n = 6-8$ .

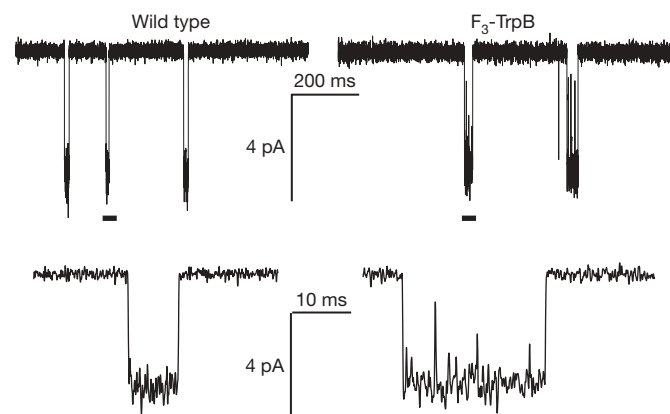
nicotine  $EC_{50}$  by a modest factor of 1.6 (ref. 25). We now find that for precisely the same change in the  $\alpha 4\beta 2$  receptor, the nicotine  $EC_{50}$  increases 19-fold, a relatively large effect for such a subtle mutation<sup>26-28</sup>. Recall that the backbone ester substitution does not destroy the hydrogen bond, it simply attenuates it. Notably, ACh, which cannot make a conventional hydrogen bond to the carbonyl, shows no shift in  $EC_{50}$  in response to this mutation (Supplementary Table 1).

This establishes that the ester mutation does not globally alter the binding/gating characteristics of the receptor.

The differential affinity of nicotine for  $\alpha 4\beta 2$  versus muscle-type receptors results from stronger interactions in the former with TrpB—both cation- $\pi$  and hydrogen bonding. Because the two receptors are identical with regard to the five residues that make up the aromatic box, a factor 'outside the box' must be influencing its precise geometry, such that nicotine can approach TrpB more closely in  $\alpha 4\beta 2$  than in muscle-type nAChR. Pioneering work has identified residues responsible for the fact that  $\alpha 4\beta 2$  receptors show consistently higher affinity than the homopentameric  $\alpha 7$  neuronal receptors<sup>29</sup>. At a particular residue in loop B—position 153, just four residues from TrpB—mutations strongly influence affinity. In high-affinity  $\alpha 4\beta 2$  receptors this residue is a Lys, and this residue is proposed to help shape the aromatic box by forming a backbone hydrogen bond between loops B and C (Fig. 1). In the lower affinity  $\alpha 7$  neuronal receptor, residue 153 is a Gly, and molecular dynamics simulations of  $\alpha 7$  suggest that a Gly at 153 discourages the formation of the hydrogen bond between loops B and C. Interestingly, the aligned residue in the muscle-type receptor is also Gly, and a naturally occurring G153S mutation is gain-of-function and associated with a congenital myasthenic syndrome<sup>30</sup>. We now report that the muscle-type  $\alpha 1$  G153K mutant shows much higher affinity for nicotine, and that, when this mutation is present, the cation- $\pi$



**Figure 4 | Fluorination plots.** Note that in both plots, all data sets share the point at  $x = 32.6 \text{ kcal mol}^{-1}$  (cation- $\pi$  energy for Trp);  $y = 0$  (black circle). Moving to the left then corresponds to monofluoro-, difluoro-, trifluoro- and tetrafluoro-TrpB. Cation- $\pi$  binding energies ( $x$  axes) are from ref. 9. **a**, Muscle-type receptor. The designation WT indicates Gly at position 153. **b**,  $\alpha 4\beta 2$  receptor.



**Figure 5 | Single-channel recordings from wild-type  $\alpha 4\beta 2$  (conventional expression) and the  $F_3$ -Trp mutant (nonsense suppression) at site B, with nicotine applied at  $EC_{50}$  values (0.080 and  $1.2 \mu\text{M}$ , respectively). Lower traces are expansions of the regions marked by a bar in the upper trace. Records were obtained in the cell-attached configuration with a pipette potential of +100 mV and are shown at 2 kHz bandwidth. Channel openings are shown as downward deflections.**

interaction to TrpB is strong. The data are summarized in Supplementary Table 1 and Fig. 4. As expected, the ACh cation- $\pi$  interaction is maintained in the muscle-type receptor with the G153K mutation. These data indicate that the loop B-loop C hydrogen bond that is naturally present in  $\alpha 4\beta 2$  shapes the aromatic box so that nicotine can make a closer contact to TrpB, and that this structural feature is absent or weaker in the muscle-type receptor.

Taken together, the present results indicate that the higher affinity of nicotine in the brain relative to the neuromuscular junction is a consequence of enhanced interactions with TrpB. A cation- $\pi$  interaction that is absent in the muscle-type receptor is quite strong in  $\alpha 4\beta 2$ . In addition, a hydrogen bond to a backbone carbonyl that is weak in the muscle type is enhanced in  $\alpha 4\beta 2$ . Both effects are quite substantial, and in combination they are more than adequate to account fully for the differential sensitivity to nicotine of the two receptors. The side chain of residue 153 in loop B distinguishes the two receptor types and apparently influences the shape of the binding site aromatic box, allowing a stronger interaction between nicotine and TrpB in high-affinity receptors.

## METHODS SUMMARY

**Whole-cell electrophysiological characterization of agonist-induced responses.** Rat  $\alpha 4\beta 2$  and mouse ( $\alpha 1$ )<sub>2</sub> $\beta 1\gamma\delta$  ion channels were expressed in *Xenopus laevis* oocytes. For  $\alpha 4\beta 2$  receptors, subunit stoichiometry was controlled by varying the  $\alpha 4:\beta 2$  subunit ratio and verified by voltage-jump experiments. Dose-response measurements for these channels were performed with a holding potential of -60 mV.

**Unnatural amino acid/ $\alpha$ -hydroxy acid incorporation.** Unnatural amino acids and  $\alpha$ -hydroxy acids were prepared, coupled to dCA and ligated to 74-mer THG73 as described previously<sup>15</sup>.

**Single-channel characterization of  $\alpha 4\beta 2$ .** Single-channel recording was performed in the cell-attached configuration with a pipette potential of +100 mV as described previously<sup>27</sup>.  $P_{\text{open}}$  values were calculated from event-detected data using Clampfit 9.2 single-channel search.

See associated *Nature Protocols* paper by X.X. *et al.* (in the press).

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 26 November 2008; accepted 9 January 2009.

Published online 1 March 2009.

- Romanelli, M. N. *et al.* Central nicotinic receptors: structure, function, ligands, and therapeutic potential. *ChemMedChem* **2**, 746–767 (2007).
- Corringer, P. J., Le Novère, N. & Changeux, J. P. Nicotinic receptors at the amino acid level. *Annu. Rev. Pharmacol. Toxicol.* **40**, 431–458 (2000).
- Coe, J. W. *et al.* Varenicline: An  $\alpha 4\beta 2$  nicotinic receptor partial agonist for smoking cessation. *J. Med. Chem.* **48**, 3474–3477 (2005).
- Gotti, C., Zoli, M. & Clementi, F. Brain nicotinic acetylcholine receptors: native subtypes and their relevance. *Trends Pharmacol. Sci.* **27**, 482–491 (2006).
- Mansvelder, H. D., Keath, J. R. & McGehee, D. S. Synaptic mechanisms underlie nicotine-induced excitability of brain reward areas. *Neuron* **33**, 905–919 (2002).
- Nashmi, R. *et al.* Chronic nicotine cell specifically upregulates functional  $\alpha 4^*$  nicotinic receptors: basis for both tolerance in midbrain and enhanced long-term potentiation in perforant path. *J. Neurosci.* **27**, 8202–8218 (2007).
- Tapper, A. R. *et al.* Nicotine activation of  $\alpha 4^*$  receptors: sufficient for reward, tolerance, and sensitization. *Science* **306**, 1029–1032 (2004).
- Kuryatov, A., Luo, J., Cooper, J. & Lindstrom, J. Nicotine acts as a pharmacological chaperone to up-regulate human  $\alpha 4\beta 2$  acetylcholine receptors. *Mol. Pharm.* **68**, 1839–1851 (2005).
- Zhong, W. *et al.* From *ab initio* quantum mechanics to molecular neurobiology: A cation- $\pi$  binding site in the nicotinic receptor. *Proc. Natl Acad. Sci. USA* **95**, 12088–12093 (1998).

- Brejci, K. *et al.* Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. *Nature* **411**, 269–276 (2001).
- Sixma, T. K. & Smit, A. B. Acetylcholine binding protein (AChBP): a secreted glial protein that provides a high-resolution model for the extracellular domain of pentameric ligand-gated ion channels. *Annu. Rev. Biophys. Biomol. Struct.* **32**, 311–334 (2003).
- Dougherty, D. A. Cys-loop neuroreceptors: structure to the rescue? *Chem. Rev.* **108**, 1642–1653 (2008).
- Beene, D. L. *et al.* Cation- $\pi$  interactions in ligand recognition by serotonergic (5-HT<sub>3A</sub>) and nicotinic acetylcholine receptors: The anomalous binding properties of nicotine. *Biochemistry* **41**, 10262–10269 (2002).
- Dougherty, D. A. Physical organic chemistry on the brain. *J. Org. Chem.* **73**, 3667–3673 (2008).
- Nowak, M. W. *et al.* *In vivo* incorporation of unnatural amino acids into ion channels in a *Xenopus* oocyte expression system. *Methods Enzymol.* **293**, 504–529 (1998).
- Fonck, C. *et al.* Novel seizure phenotype and sleep disruptions in knock-in mice with hypersensitive  $\alpha 4^*$  nicotinic receptors. *J. Neurosci.* **25**, 11396–11411 (2005).
- Filatov, G. N. & White, M. M. The role of conserved leucines in the M2 domain of the acetylcholine receptor in channel gating. *Mol. Pharmacol.* **48**, 379–384 (1995).
- Labarca, C. *et al.* Channel gating governed symmetrically by conserved leucine residues in the M2 domain of nicotinic receptors. *Nature* **376**, 514–516 (1995).
- Kearney, P. *et al.* Dose-response relations for unnatural amino acids at the agonist binding site of the nicotinic acetylcholine receptor: Tests with novel side chains and with several agonists. *Mol. Pharmacol.* **50**, 1401–1412 (1996).
- Moroni, M., Zwart, R., Sher, E., Cassels, B. K. & Bermudez, I.  $\alpha 4\beta 2$  nicotinic receptors with high and low acetylcholine sensitivity: pharmacology, stoichiometry, and sensitivity to long-term exposure to nicotine. *Mol. Pharmacol.* **70**, 755–768 (2006).
- Nelson, M. E., Kuryatov, A., Choi, C. H., Zhou, Y. & Lindstrom, J. Alternate stoichiometries of  $\alpha 4\beta 2$  nicotinic acetylcholine receptors. *Mol. Pharmacol.* **63**, 332–341 (2003).
- Celie, P. H. *et al.* Nicotine and carbamylcholine binding to nicotinic acetylcholine receptors as studied in AChBP crystal structures. *Neuron* **41**, 907–914 (2004).
- Hansen, S. B. *et al.* Structures of aplysia AChBP complexes with nicotinic agonists and antagonists reveal distinctive binding interfaces and conformations. *EMBO J.* **24**, 3635–3646 (2005).
- Talley, T. T. *et al.* Spectroscopic analysis of benzylidene anabaseine complexes with acetylcholine binding proteins as models for ligand-nicotinic receptor interactions. *Biochemistry* **45**, 8894–8902 (2006).
- Cashin, A. L., Petersson, E. J., Lester, H. A. & Dougherty, D. A. Using physical chemistry to differentiate nicotinic from cholinergic agonists at the nicotinic acetylcholine receptor. *J. Am. Chem. Soc.* **127**, 350–356 (2005).
- Deechongkit, S. *et al.* Context-dependent contributions of backbone hydrogen bonding to  $\beta$ -sheet folding energetics. *Nature* **430**, 101–105 (2004).
- England, P. M., Zhang, Y., Dougherty, D. A. & Lester, H. A. Backbone mutations in transmembrane domains of a ligand-gated ion channel: implications for the mechanism of gating. *Cell* **96**, 89–98 (1999).
- Koh, J. T., Cornish, V. W. & Schultz, P. G. An experimental approach to evaluating the role of backbone interactions in proteins using unnatural amino acid mutagenesis. *Biochemistry* **36**, 11314–11322 (1997).
- Grutter, T. *et al.* An H-bond between two residues from different loops of the acetylcholine binding site contributes to the activation mechanism of nicotinic receptors. *EMBO J.* **22**, 1990–2003 (2003).
- Sine, S. M. *et al.* Mutation of the acetylcholine receptor  $\alpha$  subunit causes a slow-channel myasthenic syndrome by enhancing agonist binding affinity. *Neuron* **15**, 229–239 (1995).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank B. N. Cohen for advice on single-channel recording and analysis. This work was supported by the NIH (NS 34407; NS 11756) and the California Tobacco-Related Disease Research Program of the University of California, grant number 16RT-0160. J.A.P.S. was partially supported by an NRSA training grant.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to D.A.D. ([dadougherty@caltech.edu](mailto:dadougherty@caltech.edu)).



## METHODS

**Whole-cell electrophysiological characterizations of the agonist-induced responses.** Rat  $\alpha 4$  and  $\beta 2$  mRNAs as well as mouse  $\alpha 1$ ,  $\beta 1$ (L9'S),  $\gamma$  and  $\delta$  mRNAs were obtained from NotI linearizations of the expression vector pAMV, followed by *in vitro* transcription using the mMessage mMachine T7 kit (Ambion). The mutations for each subunit were introduced according to the QuikChange mutagenesis protocol (Stratagene).

To express wild-type neuronal ion channels,  $\alpha 4$ L9'A mRNA was co-injected with  $\beta 2$  mRNA at various ratios (total mRNA 10–25 ng cell<sup>-1</sup>). Stage V–VI *Xenopus laevis* oocytes were injected and incubated at 18 °C for 24–48 h (whole-cell recording) or 60–90 h (single-channel recording).

Agonist-induced currents were recorded in two-electrode voltage clamp mode using the OpusXpress 6000A (Molecular Devices Axon Instruments) at a holding potential of –60 mV. Agonists were prepared in Ca<sup>2+</sup>-free ND96 solution and applied for 12 s followed by a 2 min wash with Ca<sup>2+</sup>-free ND96 solution between each agonist application. Acetylcholine chloride and (–)-nicotine tartrate were purchased from Sigma/Aldrich/RBI. Dose–response data were obtained for  $\geq 6$  concentrations of agonist and for  $\geq 5$  oocytes. Mutants with  $I_{\max}$  of  $\geq 100$  nA were defined as functional. EC<sub>50</sub> values and the Hill coefficient were calculated by fitting the dose–response relation to the Hill equation. All data are reported as mean  $\pm$  s.e.m.

Voltage jump experiments were performed in the absence of ACh and also at EC<sub>50</sub> concentration of ACh. The membrane potential was held at –60 mV, and stepped to 10 test potentials at 20-mV increments between +70 mV and –110 mV for 400 ms each. The voltage was then held for 600 ms at –60 mV holding potential between each episode. To isolate the ACh-induced currents, control traces ([ACh] = 0) were subtracted from the steady-state amplitudes of the ACh-induced currents of the test pulses. Normalized current–voltage curves were generated using current amplitudes normalized to that at –110 mV. For each  $\alpha 4$ L9'A $\beta 2$  mutant, normalized  $I_{+70\text{ mV}} \pm$  s.e.m. from  $\geq 5$  cells was reported.

**Unnatural amino acid/ $\alpha$ -hydroxy acid incorporation.** Nitroveratryloxycarbonyl (NVOC) protected cyanomethyl ester forms of unnatural amino acids and  $\alpha$ -hydroxythreonine cyanomethyl ester were synthesized, coupled to the dinucleotide dCA, and enzymatically ligated to 74-mer THG73 tRNA<sub>CUA</sub><sup>15</sup>. The unnatural amino-acid-conjugated tRNA was deprotected by photolysis immediately before co-injection with mRNA containing the UAG mutation at the site of interest. Approximately 10–25 ng mRNA and 25 ng tRNA-amino acid or tRNA-hydroxy acid were injected into stage V–VI oocytes in a total volume of 70 nl. For unnatural amino acid mutagenesis experiments in the muscle-type receptor, the  $\alpha 1$ ,  $\beta 1$ ,  $\gamma$  and  $\delta$  subunits were co-injected in a 10:1:1:1 ratio. All muscle-type receptors contained a L9'S mutation in the  $\beta$  subunit.

The fidelity of unnatural amino acid incorporation was confirmed at each site with a 'wild-type recovery' experiment and a 'read-through/reaminoacylation' test. In the wild-type recovery experiment, UAG mutant mRNA was co-injected

with tRNA charged with the amino acid that is present at this site in the wild-type protein. Generation of receptors that were indistinguishable from the wild-type protein indicated that the residue carried by the suppressor tRNA was successfully and exclusively integrated into the protein. In the 'read-through/reaminoacylation' test, the UAG mutant mRNA was introduced with (1) no tRNA, (2) tRNA THG73 that was not charged with any amino acid or (3) tRNA THG73 enzymatically ligated with dinucleotide dCA. Lack of currents in these experiments validated the reliability of the nonsense suppression experiments.

**Single-channel characterization of  $\alpha 4\beta 2$ .** Single-channel recording was performed in the cell-attached configuration on devitellinized *Xenopus laevis* oocytes at  $20 \pm 2$  °C with a pipette potential of +100 mV, as described previously<sup>27</sup>. Pipettes were fabricated from thick-walled (inner diameter = 0.80 mm, outer diameter = 1.60 mm) KG-33 glass (Garner Glass Company) and coated with sylgard (World Precision Instruments); they had resistances of 10–20 M $\Omega$ . The bath solution contained 120 mM KCl, 5 mM HEPES, 1 mM MgCl<sub>2</sub> and 2 mM CaCl<sub>2</sub>, pH = 7.4, so that the reversal potential for agonist-induced currents of devitellinized oocytes was  $\sim 0$  mV, and the transmembrane potential of the patch was  $\sim -100$  mV. The pipette solution contained 100 mM KCl, 10 mM HEPES, 1 mM MgCl<sub>2</sub>, 10 mM K<sub>2</sub>EGTA, pH = 7.4 and was supplemented with the indicated concentrations of nicotine. Data were collected using a GeneClamp 500B amplifier (Axon Instruments) at full bandwidth (50 kHz; 4-pole Bessel, –3 dB) with a CV-5 100 GU headstage. The signal was then low-pass filtered (Avens Signal Equipment, AP220, 8-pole Bessel, –3 dB 20 kHz) and sampled with a Digidata 1320A and Clampex 9.2 (Axon Instruments) at 50 kHz. Only patches that showed no simultaneous activations were analysed. For each mutant, this was  $\geq 3$  patches from oocytes from two different donor frogs. Data were filtered offline (Gaussian, –3 dB, 5 kHz) and electrical interference at harmonics of 60 Hz was removed if necessary. Event transitions were detected with Clampfit 9.2 (single-channel search). A dead time,  $t_d$ , of 100  $\mu$ s was applied to all events. Open and closed dwell time histograms were generated as described previously<sup>31</sup> and fitted using the predefined log-transformed exponential probability density function in Clampfit 9.2. To delineate clusters, a critical closed duration,  $\tau_{\text{crit}}$  was defined using two separate methods (Supplementary Discussion); in both cases closed dwell times longer than  $\tau_{\text{crit}}$  were excluded from further analysis. Sojourns to a subconductance state ( $< 85\%$  of the full conductance level) were treated as closed and accounted for  $< 10\%$  of the total openings in all records. The time-average probability that the channel is open ( $P_{\text{open}}$ ) was calculated as the total open time divided by the sum of the revised total closed time and the total open time.

31. McManus, O. B., Blatz, A. L. & Magleby, K. L. Sampling, log binning, fitting, and plotting durations of open-and-shut intervals from single channels and the effects of noise. *Pflügers Archiv-Eur. J. Physiol.* **410**, 530–553 (1987).

## CORRIGENDUM

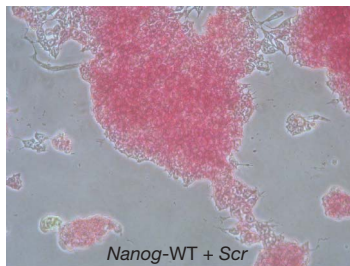
doi:10.1038/nature07880

**MicroRNAs to *Nanog*, *Oct4* and *Sox2* coding regions modulate embryonic stem cell differentiation**

Yvonne Tay, Jinqiu Zhang, Andrew M. Thomson, Bing Lim &amp; Isidore Rigoutsos

*Nature* 455, 1124–1128 (2008)

In this Letter, an incorrect micrograph was inadvertently used in the panel entitled '*Nanog*-WT + *Scr*' in row 1 of Fig. 3a: the used micrograph was from a different field of view from the '*EV* + *Scr*' experiment. A representative micrograph from the '*Nanog*-WT + *Scr*' experiment is presented below. This oversight affects neither the data nor the conclusions of the work presented in the manuscript.



## CORRIGENDUM

[doi:10.1038/nature07941](https://doi.org/10.1038/nature07941)**ABIN-1 is a ubiquitin sensor that restricts cell death and sustains embryonic development**

Shigeru Oshima, Emre E. Turer, Joseph A. Callahan, Sophia Chai, Rommel Advincula, Julio Barrera, Nataliya Shifrin, Bettina Lee, T. S. Benedict Yen, Tammy Woo, Barbara A. Malynn & Averil Ma

*Nature* 457, 906–909 (2009)

---

In this Letter, the name of T. S. Benedict Yen was incorrectly listed as Benjamin Yen. This author is affiliated with the Pathology Service, VA Medical Center, and the Department of Pathology, University of California, and not with the Department of Medicine, University of California, as originally listed.



# naturejobs

**THE CAREERS  
MAGAZINE FOR  
SCIENTISTS**

In next week's Naturejobs, readers will notice the latest step in the evolution of the section, and our most recent attempt to better serve the aspirations of career-minded scientists. Our print pages will receive a facelift that the editorial team hopes will improve the overall look, ease of navigation and our ability to provide timely information.

Along with some cosmetic tweaks will be more substantive changes. We'll be adding a regular section of news briefs covering, for example, the launch of new research centres and funded programmes, and updates on previous stories. These briefs, along with longer careers news stories, will endeavour to keep readers abreast of the current events that affect the science workforce.

The Prospects commentaries will have occasional guest contributors who will share various and different points of view from inside academia, government and the private sector. Movers will be converted into a new element called Careers Q&A, where the basic aim will be the same: to track the career arcs of science professionals in a way that informs job-seekers and others contemplating their future paths in science. But we hope Careers Q&A will be more engaging, informative and informal.

We will also include more data graphics, describing trends in employment or higher education. Sometimes a graph or table captures more clearly a growing interest — or perhaps highlights the issues at hand, whether changes in salary or the career fates of postdocs. Our Features will continue to cover a range of topics, from regional jobs hot spots to employment trends in different research disciplines. Our Postdoc Journal writers will keep documenting their aspirations and experiences in their respective fields.

These changes follow a website revamp that we launched late last year. We reorganized the site to facilitate both job searching and access to editorial content. And as always, we welcome feedback as well as story ideas. Tell us what we're doing right, what we're doing wrong, what we're missing and how we can improve.

**Gene Russo is editor of Naturejobs.**

## CONTACTS

**Editor:** Gene Russo

**Assistant editor:** Karen Kaplan  
e-mail: [naturejobseditor@nature.com](mailto:naturejobseditor@nature.com)

**European Head Office, London**  
The Macmillan Building,  
4 Crinan Street, London N1 9XW, UK  
Tel: +44 (0) 20 7843 4961  
Fax: +44 (0) 20 7843 4996  
e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

**European Sales Manager:**  
Dan Churchward (4966)  
e-mail: [d.churchward@nature.com](mailto:d.churchward@nature.com)  
**Assistant European Manager:**  
Nils Moeller (4953)

**Natureevents:**  
Ghizlaine Ababou (+44 (0) 20 7014 4015)  
e-mail: [g.ababou@nature.com](mailto:g.ababou@nature.com)

**Southwest UK/RoW:**  
Alexander Ranken (4944)

## Northeast UK/Ireland:

Matthew Ward (+44 (0) 20 7014 4059)

**France/Switzerland/Belgium:**  
Muriel Lestringuez (4994)

**Scandinavia/Spain/Portugal/Italy:**  
Evelina Rubio-Hakansson (4973)

**North Germany/The Netherlands/Eastern**

**Europe:** Kerstin Vincze (4970)

**South Germany/Austria:**

Hildi Rowland (+44 (0) 20 7014 4084)

**Advertising Production Manager:**

Stephen Russell

To send materials use London address above.

Tel: +44 (0) 20 7843 4816

Fax: +44 (0) 20 7843 4996

e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

**Naturejobs web development:** Tom Hancock

**Naturejobs online production:** Dennis Chu

**US Head Office, New York**

75 Varick Street, 9th Floor,  
New York, NY 10013-1917

Tel: +1 800 989 7718

Fax: +1 800 989 7103

e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

**US Sales Manager:** Ken Finnegan

**India**

Vikas Chawla (+91 1242881057)

e-mail: [v.chawla@nature.com](mailto:v.chawla@nature.com)

**Japan Head Office, Tokyo**

Chiyoda Building, 2-37 Ichigayatamachi,  
Shinjuku-ku, Tokyo 162-0843

Tel: +81 3 3267 8751

Fax: +81 3 3267 8746

**Asia-Pacific Sales Manager:**

Ayako Watanabe (+81 3 3267 8765)

e-mail: [a.watanabe@natureasia.com](mailto:a.watanabe@natureasia.com)

**Business Development Manager, Greater**

**China/Singapore:**

Gloria To (+852 2811 7191)

e-mail: [g.to@natureasia.com](mailto:g.to@natureasia.com)

# MOVERS

**Thomas Lynch, director, Yale Cancer Center, and physician-in-chief, Smilow Cancer Hospital, New Haven, Connecticut**



**1994–present:** Assistant to full professor of medicine, Harvard Medical School, Boston, Massachusetts

**2006–present:** Chief, haematology and oncology, Massachusetts General Hospital Cancer Center, Boston, Massachusetts

Thomas Lynch has spent 23 years at the renowned Massachusetts General Hospital (MGH) Cancer Center pushing the boundaries of personalized treatments for lung cancer, the largest cause of cancer deaths. When he becomes director of the Yale Cancer Center in April, he plans to instil the culture that yielded such success at MGH.

Yale is where Lynch received his undergraduate degree in biology, and later his MD. His proclivity for finding life-saving treatments had come from watching his father, a haematologist, struggle to treat haematologic malignancy in patients. At Yale Cancer Center, he forged an interest in translational research, including using animal models to study the mechanisms of toxicity of the cancer drug bleomycin. "I wanted to do something important in medicine, and I don't think any event is more profound than receiving a diagnosis of a dreaded disease like cancer," says Lynch.

As an oncology fellow at the Dana-Farber Cancer Institute in 1989, Lynch pursued drug treatments for lung cancer, as few treatments of value existed at the time. He helped pioneer personalized medicine, using molecular testing for genetic mutations to target lung-cancer therapies.

Happy at MGH, he had never answered enquiries from recruiters until the Yale position emerged last autumn. "Yale is one of the nation's great science cultures, yet its cancer centre is currently ranked 29th," says Lynch. But Yale administrators convinced him that the centre could foster a first-class clinical-research programme. Not only will the Smilow Cancer Hospital at Yale-New Haven Hospital open this October, but Lynch will also oversee a new institute for cancer biology at Yale's west campus.

Bruce Chabner, clinical director at MGH, is getting used to his colleagues leaving MGH to help build other centres. Lynch is the second to leave in the past year. "We've become a model for how new centres can effectively organize on a disease-specific basis across multiple departments," Chabner says. He predicts that Lynch will easily coordinate Yale's efforts.

To facilitate research in the face of an increasingly expensive regulatory environment, Lynch plans to streamline tumour-sample collection and novel tumour therapies. He says the hallmark of the Yale Cancer Center will be rapid genotyping of patients to guide therapies — the key, he says, to providing the best care possible.

"This is about recognizing the value of keeping patients at the centre of research," says Lynch. ■

**Virginia Gewin**

## NETWORKS & SUPPORT

### A mixed bag for UK universities

Administrators at the University of Cambridge are not pleased. They contend that the money they received from the UK Research Assessment Exercise (RAE), whose funding results for England were announced on 5 March, is not enough to pay for expansion of science faculty or research staff. "Is the UK going to abandon its world-class institutes?" asks Ian Leslie, the university's pro-vice chancellor for research.

The university will receive about £113.6 million (US\$159 million) in research funds next year, up 1.8% on the previous year, on the basis of its performance in the recent RAE (see *Nature* **457**, 624–625; 2009).

Leslie wonders whether the Higher Education Funding Council for England (HEFCE), which conducts the RAE, has opted to support other universities at the expense of eminent research-intensive institutions.

"These funding levels are unlikely to cause new hirings," Leslie says. "The government has made noises about maintaining the budget for higher education, but we are concerned about a move to disperse funding more widely." Under HEFCE's allocation schedule, 113 institutes will receive a share of £1.572 billion in research funding, up 7.8% from the £1.458 billion in the current year.

The University of Oxford will receive about £118.9 million, the most of any recipient and an increase of more than 7%, yet Oxford also says this isn't enough. "The increase in this funding does not keep pace with the increase in our activity or our costs," says vice-chancellor John Hood.

Nor is University College London (UCL) pleased with its £104.5 million, up 0.3%. The university has actually lost money on a per-researcher basis, says president Malcolm Grant. "So even though you increase the quality of your research, you're punished," he says. UCL will still "invest in excellence when we can", Grant says.

Not everyone is disappointed. The University of Exeter — which will get some £18 million, up 24% from last year — will fill 29 science teaching and research posts, says Roger Kain, deputy vice-chancellor of research. "We're very satisfied," he adds.

At Queen Mary, University of London, (up 29% to £32 million), and at Durham University (up 5% to £25 million), new science hires are likely. And at the University of Leicester, which will receive about £22.1 million, up 17%, Ian Postlethwaite, pro-vice-chancellor for research, says he expects to fill 20 lectureships. "This represents a real triumph," he says. ■

**Karen Kaplan**

#### POSTDOC JOURNAL

### How low can I go?

It's tiring to be in limbo as a postdoc, awaiting the right opportunity. Meanwhile, another cruel sort of limbo goes on: I feel like my lab bench is gradually lowering even as the bar for successful applicants rises higher and higher. Soon you may find me underneath it, sneaking a beer.

My latest job application did not lead to an interview. As a scientist, I hesitate to believe in fate, but my inner cheer squad shouts: "Everything's for a reason and the right job for you is just down the road!" My inner cynic then muses as to whether I took a wrong turn on that road, making the 'right job' hard to find.

So I've gained nothing — but neither have I lost anything. I'm planning to submit papers in the coming months, and to help with grant applications to ensure our lab's financial future. I'm also pursuing an independent research programme that I hope will lay the groundwork for my own science empire.

Still, contingency plans are at the fore of my frontal lobe as I note how often things fail to go as planned at the bench (PCR — ugh!). In fact, technical snags and the immediacy of grant deadlines seem to correlate tightly ( $p < 0.001$ ). With our grant madness finally past, I will find time to network with some industry contacts. I'm curious about life-sciences consultancy opportunities. Perhaps that's where I'll be able to deftly dance my way to success. ■

**Julia Boughner is a postdoc in evolutionary developmental biology at the University of Calgary, Canada.**

# String-driven thing

A knotty problem.

**Vaughan Stanger**

When it comes to generating energy, Robert Heinlein's maxim definitely applies: "There ain't no such thing as a free lunch." For now, the Sun continues to shine. If it stops any time soon, I guess you'll have a few hours, at most, to blame me.

Belinda and I first heard about extracting zero-point energy from the vacuum long before the news feeds got excited. Invariably the earliest of early adopters, we ordered a box of zero-point light bulbs as soon as we read the press release, despite the outrageous price. Set to hover, they made lovely Christmas ornaments.

Whoever said that perpetual motion doesn't work?

On the Twelfth day of Christmas, Belinda switched off the lowermost bulb and let it fall into her hand.

"Oh!" she yelped.

Worried that she might have scorched her hand, I swivelled in my chair.

"Did you burn yourself?"

Belinda served up her minted you're-such-a-dummy look. "Oh, Mike; you're not going senile on me, are you?" At which point I recalled that zero-point devices operate at ambient temperature.

"What's up then?"

"Not sure yet."

She was peering intently at what, from my vantage point, looked like nothing at all. First she bent down; then she stood on tip-toe; finally she crabbed around, bobbing her head, inspecting whatever it was, or wasn't, from every angle. Finally she gave a little nod accompanied by a grunt of satisfaction.

"Okay, take a look at the rug. From here." She indicated a point in mid-air with the tip of her index finger.

We really should have got rid of that fraying memento of our honeymoon years ago, but I peered at it as instructed. A small patch looked decidedly pale. Had I spilt something on it? No, evidently not, because when I raised my head the faded patch moved too.

"Oh, my!"

"Indeed," said Belinda.

I experimented some more. By looking through the designated spot from various directions I could make patches of wall-paper and curtains fade.

Now it was my turn to point at the flaw in space. "Is that where you hung the light, darling?"

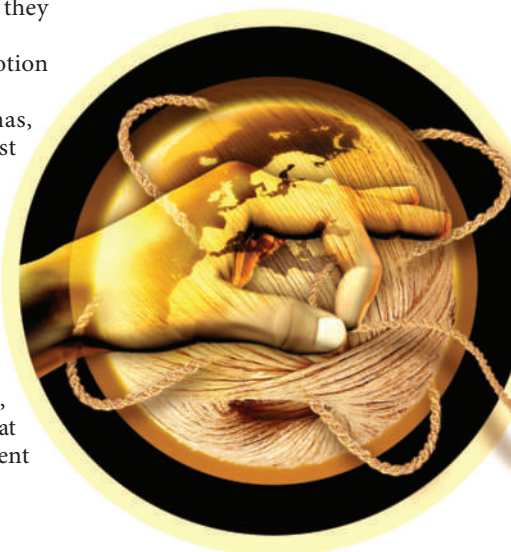
"No, that's where *you* hung it!"

I accepted the blame with a sigh before resuming my inspection of the defect. Something had caught my eye: a tiny nub of whiteness where previously there was nothing.

"Unless I'm much mistaken," I said, "there is something here now."

With my thumbnail and forefinger pinched together, I attempted to tweeze it out. On my third attempt, I finally managed to extract a short piece of what looked remarkably like string.

"Now you've done it," Belinda said.



How right she was.

That evening, I fired up my laptop and sent an e-mail to BBC News. My announcement retained 'most read' status for a fortnight. I suppose someone had to be first to report the phenomenon, which is how the Jamison Effect came to be so named.

Before long, everyone was pulling string out of nothing. A forgotten generation of cosmologists presumably jumped for joy, albeit arthritically.

A few learned voices did express concern, but most folk continued to run their vacuum energy bulbs and other zero-point devices without batting an eyelid. As for me, well, I reverted to cranking the power handle on my laptop. Scuffing up the Universe just didn't feel right, somehow.

Now, the thing about string is that it's such a versatile material, particularly useful in these resource-limited times. Belinda, bless her, decided to knit me a string vest, but I made her wear it, because, to be frank,

seeing her dressed like that did something wicked to my libido.

"Tie a knot in it, darling," she said, favouring me with a lascivious wink. I'm pretty sure she wasn't referring to the string-based necktie she'd sewn up a couple of days before, the topology of which I was still attempting to tame.

Similar scenes of blissful domesticity doubtless typified the all-too-brief era of free energy. The associated glut of string led to a cornucopia of crafts-based competitions and concerts for music played exclusively on 'stringed' instruments. Who other than me cared about the growing number of bare patches?

A killjoy physicist working for the Russian Academy of Sciences, that's who. Kudos to him, I say. He measured the speed of light every month for a year, achieving unprecedented accuracy.

His paper in *Nature* caused a sensation. Further experiments revealed that the elementary charge had increased slightly. Evidently the Universe's fundamental constants weren't so constant any more. True, the changes were tiny, but that wasn't the point. The more pieces of cosmic

string we pulled from thin air, the more deltas we inflicted on the fine structure constant. And as more than one eminent scientist has observed, the fundamental properties of our Universe do seem exquisitely fine-tuned for the existence of life. How many hits can a take before that observation no longer holds true? No one knows for sure, but I know that I never used to get this nervous before sunrise.

Only the other day, the government announced personal quotas for harvesting zero-point energy, plus a string-trading scheme. But for me, such measures don't go nearly far enough.

So, Belinda and I have bitten the bullet and declared our house a string-free zone. From now on, we'll be relying on solar power. We're the first of the un-adopters and we'll trumpet our conversion to anyone who'll listen, because the world has to learn that energy is never free.

■  
Shortly before the publication of Vaughan Stanger's story *The last botnet* (*Nature Phys.* 4, 744; 2008), NASA found some computer viruses on board the International Space Station. So, if you're reading this, the Universe may well be about to end.

JACEY